THE ELGAR COMPANION TO THE ECONOMICS OF PROPERTY RIGHTS

The Elgar Companion to the Economics of Property Rights

Edited by

Enrico Colombatto

Professor of Economics, School of Economics, University of Turin, Italy

Edward Elgar
Cheltenham, UK • Northampton, MA, USA

© Enrico Colombatto 2004

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
Northampton
Massachusetts 01060
USA

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

Library of Congress Cataloguing in Publication Data
The Elgar companion to the economics of property rights / edited by Enrico Colombatto.
p. cm.
Includes bibliographical references and index.
1. Property. 2. Right of property. I. Colombatto, Enrico.

HB701.E445 2004
330.1'7—dc22 2004041176

ISBN 1 84064 994 1 (cased)

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

Contents

Tables

Contributors

Ryan C. Amacher Professor, Department of Economics, and former President, University of Texas at Arlington, Arlington, TX, USA.

Terry L. Anderson Executive Director, Property and Environment Research Center (PERC), Bozeman, MT, and Senior Fellow, Hoover Institution, Stanford, CA, USA.

Cristiano Antonelli Professor of Economics, Department of Economics, University of Turin, Turin, Italy.

Christian Barrère Professor of Economics, University of Rheims, Rheims, France.

Norman Barry Professor and Chairman of Politics, Department of Politics, Buckingham University, Buckingham, UK.

Bruce L. Benson DeVoe Moore Distinguished Research Professor, Department of Economics, Florida State University, Tallahasse, FL, USA.

Eric Brousseau Professor of Economics, Department of Economics, University of Paris X, and Fellow, Institut Universitaire de France, and Researcher at FORUM, Centre National de la Recherche Scientifique Paris, France.

Ronald A. Cass Dean Emeritus and Melville Madison Bigelow Professor of Law, Boston University School of Law, Boston, MA, USA and Senior Fellow, International Centre for Economic Research (ICER), Turin, Italy.

Alejandro A. Chafuen President and CEO, Atlas Economic Research Foundation, Fairfax, VA, USA.

Lloyd Cohen Professor, School of Law, George Mason University, Fairfax, VA, USA.

Enrico Colombatto Professor of Economics, School of Economics, University of Turin, and International Centre for Economic Research (ICER), Turin, Italy.

Tiziana Cuccia Associate Professor of Economics, Department of Economics, University of Catania, Catania, Italy.

Pierre Garello Professor, Department of Economics, Center for d'Analyse Economique, University Paul Cézanne Aix-Marseille III, France.

Pierre Garrouste Professor of Economics, Department of Economics, University Lumière-Lyon 2 and ATOM, University of Paris 1, France.

Hans-Hermann Hoppe Professor of Economics, Department of Economics, University of Nevada, Las Vegas, NV, USA.

Alan G. Isaac Associate Professor, Department of Economics, American University, Washington, DC, USA.

Dwight R. Lee Professor of Economics, Department of Economics, University of Georgia, Athens, GA, USA.

Gary D. Libecap Professor of Economics and Law, University of Arizona, Tucson, AZ, USA, Research Associate, National Bureau of Economic Research, Cambridge, MA.

Leonard P. Liggio Research Professor, George Mason University School of Law, Executive Vice President, Atlas Economic Research Foundation, and Distinguished Senior Scholar, Institute for Humane Studies, George Mason University, Fairfax, VA, USA.

Fred S. McChesney Class of 1967/James B. Haddad Professor of Law, and Professor, Department of Management and Strategy, Kellogg School of Management, Northwestern University, Evanston, IL, USA.

Bobby McCormick Professor of Economics and BB&T Scholar, John Walker Department of Economics, Clemson University, Clemson, SC, USA, and Senior Associate, Property and Environment Research Center (PERC), Bozeman, MT, USA.

Roger E. Meiners Professor of Economics and Law, Department of Economics, University of Texas at Arlington, Arlington, TX, USA.

Seth W. Norton Norris A. Aldeen Professor of Business, Department of Business and Economics, Wheaton College, Wheaton, Illinois, USA.

Francesco Parisi Professor of Law and Director, Law and Economics Program, George Mason University, Fairfax, VA, USA.

Walter G. Park Associate Professor, Department of Economics, American University, Washington, DC, USA.

Steve Pejovich Professor Emeritus, [dept], Texas A&M University, College Station, TX, USA, and Senior Fellow, International Centre for Economic Research (ICER), Turin, Italy.

Andrzej Rapaczynski Daniel G. Ross Professor of Law, Columbia University, New York, NY, USA.

Walter Santagata Professor of Economics, Department of Economics, University of Turin, Turin, Italy.

Stefan Voigt Professor of Economics, Economics Department, Kassel University, Kassel, Germany, and International Centre for Economic Research (ICER), Turin, Italy.

Foreword

*Steve Pejovich**

From the very beginning of recorded history, people have understood the importance of property rights for their survival. In a very illuminating article on 'Aristotle on property rights', Fred Miller[1] traces the recognition of the social consequences of property rights back to ancient Greece. However, most mainstream neoclassical economists considered private property rights as an exogenous constraint.

It was only in the 1960s that systematic effort to internalize changes in property rights into a theoretical framework began in earnest. Within a few decades a number of scholars including Armen Alchian, Harold Demsetz, Henry Manne, Douglass North, Richard Posner and Oliver Williamson produced a significant body of literature on the relationship between property rights and economic performance. On the one hand, their research demonstrated that different property rights have specific and predictable effects on economic behavior. On the other, they were able to show that the interaction between individuals' search for more knowledge and resulting changes in the economic conditions of life affect the creation and/or modifications of property rights. Of course, by demonstrating that systematic two-way relations exist between alternative property rights and economic behavior, those scholars did much more than merely explain the effects of alternative property rights on economic behavior. They created the economic theory of property rights.

By the early 2000s, the economics of property rights has scored impressive academic gains. Recognizing those gains, Gregory Alexander (2003, p. 734) said: 'If its war against bird lovers, tree huggers, and other like minded collectivists is not yet entirely won, at least the pendulum seems to have swung in favor of the [property rights] movement'.[2] Moreover, annual publications such as *The Index of Economic Freedom* published jointly by the Heritage Foundation and the Wall Street Journal, and the *Economic Freedom of the World* published by the Fraser Institute have illuminated a strong positive relationship between private property rights and economic performance. Finally, economic reforms in Central and Eastern Europe have shown that it is not enough to just create private ownership. To close the gap between private and social costs of economic activities, it is necessary to ensure that private property rights are stable and credible. In other words, the rule of law must come before democracy.

In 22 well-written chapters, this volume illuminates the major accomplishments of the economics of property rights. While they differ in the scope and extent of their coverage, contributions to this book are focused, carefully researched, well argued, and readable. In my judgment, the book takes us to the frontier of the growing stock of knowledge on the origins and consequences of alternative property rights.

Notes

* Professor Emeritus, Texas A&M University College Station, Texas, USA, and Senior Fellow, International Centre for Economic Research, Turin, Italy.

1. Fred Miller (1991), 'Aristotle on property rights', in J. Anton and A. Preus (eds), *Essays in Ancient Greek Philosophy*, New York: State University Press of New York, pp. 227–61.
2. G. Alexander (2003), 'Property as a fundamental constitutional right? The German Example', *Cornell Law Review*, **88** (3), 733–78.

Introduction

Enrico Colombatto

On property rights and economic analysis

The role of property rights in economics can hardly be overemphasised. Indeed, since all the essential controversies about the purpose and nature of economic science can be reformulated as debates on the features of individual behaviour (consumption, production, leisure, policy making), and since the motivation and opportunities for human action are defined by systems of property rights, it is fair to say that economic analysis is about the consequences of the assignment and use of property. Of course, economics still remains a question of scarcity, as individuals try to enhance their well-being in a world characterised by ignorance and limited resources. But the very concept of scarcity would make little sense unless the idea of property were taken into account, for property rights define the structure of incentives that lead agents to struggle, compete and also cooperate in order to satisfy their needs and ambitions. Indeed, property rights characterise scarcity itself, which becomes a meaningful concept only if the possibility of using – or benefiting from – a given good is related to how much one gains by transferring (acquiring) those rights to (from) others.

This view is now commonly accepted; even by those who favour technocratic approaches based on formal models where institutions hardly play a role. Contrary to common belief, however, the link between economic performance and property rights goes well beyond the standard problems of static optimisation, whereby supposedly efficient techniques are necessarily adopted and inputs fully employed (whatever these terms mean). In fact, the major impact of property rights economics concerns ethical and dynamic issues. This is actually the focus of the chapters presented in this volume, which examine the effects of property rights (or lack of) from three different viewpoints: ethics, incentives and efficiency. In turn, these elements justify the origin and contribute to moulding the dynamics of institutions following changes in ideology, in technology and in the rules of the game typical of the law-making process.

Ethics

As aired above, and despite frequent statements generated by scientism-prone scholars, economic reasoning cannot be amoral. The notion of right and wrong plays an essential role in shaping human behaviour. More pre-

cisely, all economic actions refer to how people exploit resources either by making use of widely accepted (legitimate) means, or by violating some behavioural pattern and therefore acting illegally or illegitimately – that is against formal or informal rules, respectively.[1] Put differently, the shared notion of property rights also defines the borderline between legitimacy and crime. For instance, using violence to cause damage to, or take away resources from, an individual (nuisance and theft, respectively) may be a crime under some circumstances, but may turn out to be acceptable in other situations, say when one of the parties involved is the state.

This is why the economics of property rights is not merely a methodological question, but first of all – and perhaps foremost – a moral issue (see in particular Liggio and Chafuen, Chapter 1). The ethical features of property define its customary origins, and thus to what extent individual or collective properties can be infringed upon and the rights attached to property can be modified. They also explain how property rights evolve. In the end, both the extent to which rights can be claimed and the protection they are awarded become an ideological issue.[2]

Surely, ideology is not enough. Exogenous historical events and technological shocks also play a role, as several chapters in Part III of this volume document at length. But one should not look much further. In particular, one should think twice before accepting *in toto* the neo-institutional standpoint, whereby property rights are supposed to evolve in some kind of path-dependent fashion, with little hope for individuals and policy makers alike to change the course of events or to see them change through exogenous pressure. Not only would it be a questionable exercise in historicism, therefore with modest explanatory power.[3] It would also convey the false impression whereby our immediate future depends on built-in, impossible-to-reverse mechanisms inherited from the past. In a nutshell, individual self-responsibility *vis-à-vis* institutions and institutional change would be seriously weakened. Although the contributions to this volume reflect different moral views, all of them stay clear of the deterministic trap. As Liggio and Chafuen note at the end of their chapter, we are all responsible for the way property rights are going to develop in the future. Changing outcomes by decree is impossible. Efforts in that direction – and ultimately central planning – may lead to tragic results. Instead, our best defence is to identify the role of property rights, and realise that making them subject to discretionary power may have far-reaching and not always well-understood consequences.

Efficiency, externalities and property rights dynamics

By and large, in the recent past the danger of legal constructivism is gradually being appreciated and the necessity of limiting state intervention recognised. Still, it is also claimed that the pervasive presence of alleged

market imperfections weakens the case for unconstrained property rights. In other words, transaction costs often prevent their satisfactory specification and enforcement.[4] True enough, assessing the boundary of property rights always implies some costs, especially if they relate to new goods or services, possibly produced by means of previously unassigned resources. Acquiring the information necessary to exchange property rights (contracts) is also expensive, especially in large communities where contracts are not necessarily the repetition of past transactions concerning the same goods, or carried out by the same people with constant incentive structures. Similar comments apply to property rights enforcement. More generally, it is impossible to avoid that an action carried out by A might actually interfere with B's alleged property rights. And it may also happen that A reduces B's happiness even if no infringement of B's property occurs. Envy is a frequent cause for this. One can imagine many other examples: for instance, competitive action by company A may negatively influence B's profits, or new consumers may compete for the same product and thus provoke a rise in prices, thereby reducing somebody else's surplus (see also Hoppe, Chapter 2).

But the very recurrence of externalities explains why exercises targeted at conceiving the perfect property rights structure tend to be in vain, even if individuals had unlimited abilities to produce, store and elaborate information. Of course, such exercises could also be harmful. Property rights systems should evolve according to the relative value of the resources and goods involved, as well as to the technical possibility of enforcing such rights. At the same time, efforts to develop adequate enforcement technologies are bound to intensify as the expected benefits from enhanced monitoring increase. By looking for perfect arrangements one actually takes for granted what is in fact still unknown. As mentioned earlier, central planning often seems to be an attractive solution; but it is a deadly embrace.

In short, it is certainly justified to point out the range of externalities that are associated with most – if not all – property rights systems originating from spontaneous interaction and voluntary agreement. Many chapters in this volume provide plenty of examples along these lines. Still, one should be very cautious when replacing unconstrained rights of property with alternative regimes, whereby rights are assigned or reassigned at the discretion of some more or less legitimised ruler through top-down processes, allegedly in order to reduce externalities. It can be easily shown that when this happens, new sets of externalities tend to replace the original ones. Moreover, privileges are created and arbitrarily distributed. Indeed, history shows that private interactions and entrepreneurship are generally far more effective in reducing externalities than social engineers designing 'better' or 'fairer' property rights systems.

An introduction to the *Companion*

The questions related to the origin, evolution and effectiveness of economic systems founded on property rights regimes are analysed and assessed in the three parts of this volume. Part I deals with the foundations and thus the legitimacy of property. The boundaries of private property and the extent of legitimate or tolerable interference by policy makers are discussed in detail. Part II is devoted to the analysis of different legal systems from a property rights standpoint and their implications as regards the economics of institutions. Finally, Part III provides some applications of property rights economics to the real world. In particular, it offers a number of important regulatory challenges, and implicitly questions the ambition to achieve static efficiency and 'equitable' distribution, possibly replacing legitimacy with legality.

The origins of property rights

The first chapter examines the cultural roots of the notion of property and of private property in particular. As Liggio and Chafuen (Chapter 1) document in detail, this idea was already prominent at the dawn of the so-called 'Western civilisation'. Indeed, it had very substantial religious connotations in the Classical world, where private property was sacred because divine ancestors protected it. Violating property was thus equivalent to sacrilege.

But property was by no means a feature limited to the pagans. The authors are extremely persuasive in describing the role of private property both in the Old Testament and in the Christian history of thought as we know it from Saint Paul, Saint Augustine and Aquinas, to name just a few 'giants' of Western thought. As for the future, Liggio and Chafuen observe that attacks have been frequent over the centuries and on both sides of the Atlantic. Until recently these were successfully resisted because – the authors conclude – morality was dominated by deep individual religious awareness and commitments. The strength and quality of such ethical commitments will thus shape the reaction to the current and forthcoming threats.

Hoppe (Chapter 2) develops the discussion of the moral element in a private property rights system, and argues that unconstrained property represents the only feasible solution to the problem of social order in a world of scarcity. Contrary to what is usually claimed, a world without property rights would not only be one without much peaceful interaction (the problem of social order in its traditional form), but also one where individual behaviour would not be legitimate. For human behaviour can only take place if actors own their body and have a claim on the space they are occupying to start with. This and the Pareto optimality criterion for social welfare justify the rule of 'originally appropriated places and goods'. This means that one may appropriate resources (i) if they have not been previously appropriated by

others, which also implies that B cannot be hurt by A's appropriation, otherwise B would have established his/her property rights before A; *and* (ii) if the owner to be demonstrates that these resources are enhanced by his/her own work.[5] In all other cases resources can be appropriated only through voluntary exchange, otherwise it would be a violation of somebody's established property rights, that is, a violation of the basic rule of social order.

By calling upon the history of Western thought extending from Locke (at least) to Rothbard, Hoppe shows that there are only two alternatives to action and interaction based on property rights. One is an unequal world populated by *Uebermenschen* (who decide and are allowed to act) and *Untermenschen* (who take orders and have to ask permission for whatever they do). In other words, one should assume that not all men have equal dignity. The other alternative is 'universal communism', which is of course unfeasible, for everybody should ask permission from all other individuals (including those yet to be born), lest externalities are created.

Although the moral foundations of property rights were clearly perceived both in the Classical world (Aristotle and then the Roman legal tradition) and in the Western (Christian) civilisation, according to Hoppe such principles have been substantially weakened in recent times, as economics became increasingly involved with providing answers to political needs, while political philosophy tried to provide new alibis to policy makers. Sometimes these efforts were transparent. Sometimes they were the indirect results of well-meaning research programmes. Among the latter, during the past decades several Chicago scholars deserve credit for having drawn attention to the property rights problem from the standpoint of economic efficiency. But this should not conceal the fact that these authors ended up by denying the moral dimension of economics and allowed the initial property rights assignment to be overruled following the onset of a technological environment. Hence, the apparent legitimacy of technocratic conjectures, of cost–benefit analyses based on social utility maximisation, and of the gradual transformation of property rights into a matter of technocratic discretion.

Parisi (Chapter 3) does not deny the moral dimension of property rights structures, but tends to devote more attention to their dynamics as a path-dependent process, hopefully with a happy end. In particular he observes that as a result of demographic pressure and/or new social organisations – the family replacing the clan – collective-property systems gradually weakened. Private rights proved to be an effective response to the scarcity problem as they enhanced a more efficient use of increasingly scarce resources. Similar remarks apply to the notion of absolute ownership, introduced within the context of Roman law and further enhanced after the feudal period. Indeed, after the French Revolution, legislators enforced such principles even against the will of the parties. However, the tensions that followed ultimately led to

more freedom of contract in the twentieth century, in both civil and common law jurisdictions.

There is no doubt that property rights are the key to economic growth and enhanced well-being. Nevertheless, Parisi's conclusions are not warranted. Today much of the world's population live in societies where property rights are badly specified and poorly enforced. Indeed, property rights have failed to evolve in desirable ways. One needs to clarify why.

The contributions by Norton (Chapter 4) and Libecap (Chapter 5) offer two different but mutually compatible accounts. On the one hand, Norton calls the readers' attention to cultural elements. After carefully reviewing the debate on the role of culture as an explanatory variable for aggregate economic performance, he focuses on the link between culture and institutions, the former being understood as a 'system of values and beliefs'. In particular, by examining data over the 1982–95 period, Norton finds that only individualism is strongly correlated with a satisfactory property rights system, while ethnic homogeneity, religion and path dependence seem to be playing a minor role. He concludes that since ethical systems are hard to transmit and absorb in the short run, relatively quick, top-down transfers of successful institutional systems to relatively poor countries are virtually impossible.

On the other hand, Libecap underscores transaction costs as the major impediment to moving from one property rights system to another. For instance, lack of information prevents economic actors from appreciating the benefits of institutional change, collective-action problems make negotiations difficult, while some actors – including politicians – decline to consider a more efficient distribution of rights if it would jeopardise their rents (privileges) and powers. In all these cases, institutional change slows down or stands still. Two historical examples persuasively document the nature and magnitude of such barriers to change: the American Great Plains in the late nineteenth and early twentieth centuries, when the climate changed significantly, and the reallocation of water rights in semi-arid western US states. In the former case, efficiency would have required larger plots, but politicians opposed the concentration of property fearing migration out of the region and thus a smaller electorate, which in turn would have diluted political power at the federal level. Of course, compensation for these lost rents could not be proposed for lack of moral legitimacy. Thus, when crisis struck, subsidies were preferred to superior institutional solutions, which would have allowed faster farm consolidation and the adoption of better farming technologies.

The second example illustrates the difficulties met in reallocating water from relatively inefficient agricultural to more desirable urban uses. In particular, rights could not be transferred due to high negotiation costs (heterogeneous and fragmented owners or jurisdictions) and inadequate information. Legislative efforts to provide top-down rules of the game –

sometimes also driven by equity concerns and electoral interests – added further ambiguities and uncertainty.

Property rights and the law

All the contributions presented in Part I maintain that property rights are a fundamental condition for growth. But that condition is not sufficient when the moral foundations of property are undermined, or transaction costs are kept too high for beneficial evolution to occur. Having this framework in mind, Part II offers the basic elements for understanding how and to what extent policy makers (including the judiciary) try to influence and/or create property rights structures.

At one extreme, Barrère (Chapter 6) discusses the notion of legal efficiency by assigning the judiciary the role of an *Uebermensch*, in Hoppe's terminology. From a different perspective, Voigt (Chapter 7) and Barry (Chapter 8) deal with the issues raised by the comparison of different legal regimes. Finally, Rapaczynski (Chapter 9) and Cass (Chapter 10) evaluate whether an effective property rights system does require constitutional protection against possible attacks by populist policy making, so as to provide and enforce the rule of law.

Barrère does not deny that a clear assignment of property rights and a strict and often formalised system of sanctions is required to obtain economic growth. He nevertheless believes that the policy maker and the judge have a moral duty to exercise discretion for two sets of reasons: first, in order to lower transaction costs (contracts can never be complete), and accelerate trial-and-error processes within a 'Hayekian' society; and second, he claims that since all legal approaches to property necessarily affect individual well-being, the judge has a duty to conceive and enforce viable compromises when property rights are not assigned fairly or clearly enough. In other words, whenever the market process is supposed to need help.

Voigt is less enthusiastic about top-down solutions by more or less omniscient policy makers that know what is best (or will be best) for society. Indeed, in his view the very fact that different groups may prefer different sets of rules is a powerful source of potential conflict.[6] This happens, for instance, as a consequence of historical and political accidents (political unification through the rise of the nation-state, colonisation, or transition from communism), so that the rules of a new regime partially or totally supplant those typical of the previous one. It can also occur in heterogeneous societies, where different groups try to promote their own interests, possibly at the expense of others.

Voigt argues that under such circumstances the persistence of unofficial systems remains pervasive, as witnessed by the existence of large underground activities, including corruption. The cost of having one or more

systems using violence to suppress rival arrangements may be considerable, since individuals choosing one system must suffer the cost of escaping sanctions from the others. Hence, there is a risk that more resources will be spent by each of the various authorities to enforce their own preferences.[7]

Since the legal present of a country depends on its legal history (path dependence), according to Voigt there is little one can do to make sure that relatively peaceful competition prevails upon violent conflict, unless legal transplants are attempted, and new path-dependent processes set in motion. Still, the author aptly raises a number of doubts about the possibility of transferring property rights systems from one country to another. Perhaps transfers *de iure* are technically feasible. But of course, history demonstrates that enforcing procedures is not enough to warrant success, and that top-down attempts to introduce supranational legislation may be part of the problem, rather than of the solution.

Indeed, it is not even quite clear what kind of system should be transferred for the sake of efficiency. For instance, the alleged superiority of the common law system cannot be taken for granted. Barry strengthens and deepens Voigt's insights on this.

In particular, Barry's chapter compares procedures under the civil and common law systems. The former presents itself as a set of rules originally rooted in fundamental principles (natural order), to which rules and regulations – and also different court systems – have been added in order to accommodate new needs. On the other hand, the common law system is based on precedent (past experience being a guide to the actual will of the contracting parties), with the judge in charge of finding out the relevant precedent and of applying it. By not being anchored to a system of fundamental principles, but rather to a criterion of 'acceptable behaviour', the common law system is thought to be simpler, more flexible, open to innovative decisions and more conducive to voluntary agreements between the parties involved. That is why common law is usually thought to be closer to the needs of a modern market economy.

Despite apparently serious diversities, however, Barry claims that neither system has in fact been able to resist pressure towards regulation. That is, although significant differences remain, for instance when assessing property rights in the presence of externalities, both standards have been suffering from political influence, special interest groups and government intervention. Similarly, takings with compensation seems to be the prevailing utilitarian solution under both legal systems, even if that is an obvious breach of property rights principles and even when state authorities are not involved. In other words, Barry sees the divergence between common and civil law systems become increasingly narrow as a consequence of state intervention. It is therefore hard to claim that one system is better than others at protecting

property rights. More relevant alternatives are perhaps to be found in constitutions, some of which may well be more effective than others.

In this respect both Rapaczynski and Cass have doubts, though. In particular, Rapaczynski reviews some key elements of constitutional theory: the so-called 'personality theory'[8] is too vague to justify the very existence of property rights, let alone warrant constitutional protection. Protecting and enforcing property rights in order to support political freedom and ultimately assert the notion of citizenship is a more persuasive argument. If so, can one then claim that constitutions are a suitable device to enhance such protection and enforcement?

Rapaczynski does not think so. He maintains that constitutions are not designed to protect unrestrained economic freedom. Thus, they should not be used to that purpose. For example, the American experience proves that constitutional law making has been far more successful at shielding basic political freedom than at stopping political interference (including regulation) in private economic activities. According to the author's viewpoint, however, this may not necessarily be undesirable, for constitutions should always be subdued to the principles of democratic rule. As a matter of fact, by adopting a rather cautious attitude *vis-à-vis* classical and radical liberalism he argues that constitutions do not and should not protect wealth creation in the face of democratic policy making, especially when redistribution of wealth is in order. Furthermore, there are other good reasons not to provide unconditional protection (including full compensation for takings) to private property rights, for instance, when some private rights are harmful to what Rapaczynski identifies as 'social efficiency'.

Other tools may be more suitable in preserving private property rights, if and when it is appropriate to do so. To this purpose, Rapaczynski favours the adoption of schemes that make the goals and outcomes of policy action more transparent, and suggests that private property rights can indeed be legitimately weakened or simply violated if that enhances social efficiency. This is why well-designed and easily understandable evaluation procedures – perhaps subject to frequent electoral tests – are to be preferred to the rigidities of a constitutional norm.

On the other hand, Cass considers that the features of societies that deviate from the rule of law – including the right to engage in voluntary contracts – are weak. Similarly to Rapaczynski, Cass does not deny that property rights systems are subject to change and evolution. Contrary to the previous author, however, Cass concentrates on how change comes about and on how effectively discretionary power is constrained, which is the very purpose of constitutional safeguards. Indeed, Cass posits that legal systems and economic performances differ across countries as a consequence of the quality of the constraints imposed upon official discretionary powers. For instance,

Western democracies have been far more successful in limiting arbitrary violations of property rights than autocracies in other parts of the world.

In this light, Cass observes that the ability of a system to limit discretion is not necessarily embedded in the formal rules (constitutions). To clarify this point, the author compares the cases of Zimbabwe and the United States. In the former country, constitutions did establish sets of rules, but the ruling élites never found substantial difficulties in changing the constitution at will. As a consequence, the law was not broken, since constitutional principles could be amended legally, despite occasional opposition by the Supreme Court. But rule of law was of course shattered.

On the contrary, US procedures can do much more to protect property, mainly thanks to the system of checks and balances among the different jurisdictions, which is a typical feature of the American structure of power. As a result, changing the law becomes a slow and often difficult affair. It adds certainty and stability to the rules of the game.

The last part of the chapter discusses whether there can be desirable exceptions to the rule of law, that is, whether discretion is to be allowed under given circumstances. Cass's answer is positive, as long as individuals can take decisions in a predictable enough environment. As a matter of fact, he believes that predictability turns out to be a better principle than discretion when evaluating the quality of a property rights system and the legitimacy of changes imposed by the authorities. Discretion is necessary whenever the norm has some degree of generality – which is inevitable, since the legislator cannot be omniscient and monitoring costs are far from zero. From this standpoint, therefore, the difference between Zimbabwe and the United States is not discretion *per se*, but the fact that while discretion is constrained by cross-checks in the United States, it is virtually unbridled in Mugabe's regime.

Property rights economics in action

Most of the theoretical problems raised in Part II reveal their implications when property rights principles are applied to real life. Some authors seem to be closer to the views held by Hoppe and Norton, who argue that property rights are rooted in the natural order and therefore can never be violated. When such natural rights are tampered with, following cultural and ideological fashions, economic performance will suffer. Most efforts should thus be devoted to enhancing the definition and enforcement of property rights – hopefully in a cultural environment favourable to private property rights structures. On the other hand, some scholars are more inclined to allow weakened property rights systems so as to overcome unquestionable inefficiencies provoked – mainly – by collective-action problems. Part III of the *Companion* is devoted to these differences.

Colombatto (Chapter 11) belongs to the first group. He draws attention to the cultural traits of a society – morality once again – in order to cast doubts upon the recent literature on development economics, which has moved from neoclassical policy making, towards institutional constructivism, and now advocates the creation of allegedly suitable property rights structures to promote growth in poor countries. The author claims that the enforcement of property rights is surely crucial for development, but that it is a mistake to believe that property rights can be used as a tool for policy making and social engineering, following the latest Western tradition and experience of weakening private property. By doing so, Western developed countries have actually undermined the case for a social order based on self-responsibility. Entrepreneurship has been discouraged and the opportunities for growth have been further cut back. In short, it is hard to see how the poor can improve their well-being by referring to a moral system that keeps undermining self-responsibility and entrepreneurship.

Cohen (Chapter 12) also refers to the moral foundations of property rights in order to develop a set of powerful arguments to justify human genes manipulation (germline engineering). He substantiates his claim by referring to (i) the little disputed property right of each individual over his/her own body; a right that can be legitimately violated only when individuals pose a serious threat to other human beings. Furthermore, he recalls that (ii) in our civilisation parents can legitimately decide on behalf of their children – or of their unborn children – until their offspring can express their preferences and be aware of the possible consequences. Surely, if parents are assumed to act in the interest of their offspring outside the domain of germline engineering, then there is no reason to believe that they would change their attitude once genetic manipulation is admitted. Consistent with the free-market tradition, Cohen does not claim that parents always make the best possible choices. Nobody is perfect. But he does not believe that on average a bureaucrat or a policy maker can do better when deciding on behalf of somebody else's children.

There is no doubt that germline engineering may be a source of externalities. Most human choices are, as mentioned earlier. But as long as human action does not use violence in order to inhibit other people's choice or to reduce the right to enjoy property,[9] one has no right to restrain behaviour. This is again a moral problem which, however, leads to dynamic efficiency. The history of humankind shows that extensive choice and non-violent competition for scarce resources have been crucial in reducing rents and enhancing progress; even if gains have seldom been spread evenly across the population.[10]

In the chapters that follow, the focus shifts from the moral implications of property rights economics to the problems due to scarcity and externalities,

the latter being interpreted as the failure to develop a suitable set of contracts. Consistent with this view, Anderson and McCormick (Chapter 13) concentrate on the environment and the so-called divergences between private and social outcomes, which is in fact another way of describing the problem of conflicting uses for a scarce resource when property rights are not well defined. In particular they apply Coase's seminal idea on the nature of the firm, whereby previously abundant resources are transformed into valuable assets through contracts, thereby specifying (transferable) property rights and eliminating the tragedy of the commons. Thus, the chapter argues that entrepreneurship in an environmental context consists of the ability to perceive scarcity before others do, and find ways to overcome contracting costs.

Of course, Anderson and McCormick do not belittle the cost of monitoring and transferring property rights. Still, although voluntary exchange through compensation is not always legally necessary, potential regulators should bear in mind that top-down directives often lead to resistance and thus to inefficiency. This is indeed similar to what happens in the firm, where motivation – rewards and penalties for good and bad decisions and appropriate reactions to new information – works better than simple rule making. In short, Anderson and McCormick claim that the celebrated problem of externalities is not an issue, but the consequence of an entrepreneurial failure to reduce transaction costs and develop suitable contractual arrangements. Hence, more efforts should be devoted to enhancing entrepreneurship as a way to solve scarcities in environmental goods, rather than taking externalities for granted and designing optimal policies to regulate and constrain economic activities.

The vulnerability of property rights economics *vis-à-vis* regulation is the subject of Lee's contribution (Chapter 14), where it is maintained that the crucial role of property rights and voluntary exchange within a market system is seldom perceived, let alone acknowledged. This is due to ignorance, avarice and envy, as well to the dispersion of benefits and the concentration of costs. Similarly, it is also overlooked that when property rights are not defined, enforced and exchanged, individuals are no longer responsible for their actions. Hence, freedom is more likely to be violated. Not surprisingly, regulation and coercion tend to be the suggested solutions to redress alleged unfairness and obtain political consensus,[11] not freedom. To make his point, Lee recalls the essence of the public choice literature, which emphasises the high cost the general public incur to acquire information about the damages generated by coercive policies; while the beneficiaries need little encouragement to perceive the gains from rent seeking and forcefully engage in rhetoric exercises to persuade public opinion.

The remaining part of his contribution provides three fitting examples. One discusses the case of mandatory benefits for employees, to be provided by

their employers. Such benefits are generally perceived as a protection for the worker, whereas they are in fact equivalent to forcing workers to buy services they are not willing to buy (otherwise no compulsion would be necessary). More important, Lee shows that the true beneficiaries are indeed the inefficient producers, who try to raise costs for their competitors, and the workers' cartels, which aim at raising entry barriers for low-cost (unskilled) labour.

The second example focuses on the restrictions imposed on hostile takeovers. Economics has made clear that hostile takeovers are the most powerful way to restrain potential misbehaviour by incumbent managers. Misconduct would reduce the value of the shares, make the company vulnerable to a takeover by new owners who then replace at least part of the management. In other words, hostile takeovers are good for stockowners, and harmful for bad managers. Not surprisingly, the latter are most active in mobilising public opinion so as to raise the cost of hostile takeovers and – more generally – restrain competition. While they claim they are fighting to avoid shutdowns and losses of employment, they are in fact protecting their own jobs against the public interest.

Along similar lines, the case for global warming shows that the major beneficiaries from the regulatory policies to reduce temperatures is not the world population, but the huge bureaucracies that would be created to run these world-scale programmes. Frightening the population in order to justify such programmes is of course part of the game.

Two among Lee's insights are further developed in the next two chapters. Benson and McChesney (Chapter 15) devote their attention to corruption: an important aspect of bureaucratic organisations and – more generally – of societies where regulation has replaced property rights. Garello (Chapter 16), on the other hand, examines the domain of corporate governance.

In a nutshell, Benson and McChesney follow Lee's approach in that they consider the regulatory solution generates an adequate property rights regime, and discover that corruption in the public sector is the byproduct of such a solution. The economic profession usually regards corruption as something morally deplorable and a source of deadweight losses. On the other hand, according to Benson and McChesney, public corruption originates from the property rights attributed to government officials, so that bureaucrats and politicians end up by enjoying considerable coercive power – for instance, as regards taxation – but are hardly accountable for their conduct. This explains corrupt behaviour not only under the form of embezzlement, but also – and more appropriately – when officials are bribed not to enforce the law, to pass or not to pass legislation (rent seeking and rent extraction, respectively).

This is why corruption can be conceived as a form of payment to the policy maker in order to reduce his/her discretionary power. And discretionary or regulatory power is of course a matter of property rights. In the end, the

economics of property rights helps once more to clarify that regulation is part of the problem, rather than of the solution. Hence, as the authors argue at great length, little can be expected from the efforts to stamp out corruption unless one radically changes the institutional framework (that is, property rights assignments).

Within the areas of shareholders' and managers' behaviour, and similarly to Lee, Garello attributes the ease with which regulation is enforced to ignorance as regards the origins of the modern corporate firm and of limited liability in particular. Limited liability emerged as a contractual arrangement designed to raise large amounts of capital when owners' control necessarily becomes weaker. As a result, exit costs for the owner are reduced and part of the risk is shifted to lenders. Two cases are illustrated to clarify the point: legislation on insider trading and on takeovers.

Indeed, by regulating 'insider trading' in order to make sure that owners have equal access to information means that one has not really understood the role of the owners in a limited-liability company. Under such circumstances, shareholders have explicitly decided not to incur the cost of acquiring all the information necessary to strictly monitor how the company operates. In other words, regulating insider trading cannot be explained by the need to protect a property right on information – and thus on potential profits – held by the shareholders, for the shareholders of a limited liability company clearly chose not to gather information. Surely, if deemed desirable, shareholders may decide to force managers to abstain from trading company shares.[12] But of course, that is just a voluntary contract among the majority of shareholders, with no need for outside (state) regulation.

Legislation against takeovers is also the result of ignorance, for takeovers usually benefit the shareholders of the target company without harming the buyer. Indeed, it is a situation whereby outsiders take care of the monitoring that shareholders find too expensive to carry out themselves. Once again, this is revealed by the very essence of a limited liability contract.

Garello thus concludes that in both cases – insider trading and takeovers – the true beneficiaries from regulatory legislation are not the shareholders, but rather easy to identify pressure groups. Support for normative intervention will then be found among professional providers of information when it comes to insider trading, rent-seeking managers when it comes to hostile takeovers, and all those who have an interest in stopping competitors from becoming stronger when it comes to mergers (including those accomplished by means of friendly takeovers). In addition, one should not forget the role of alleged regulatory experts and bureaucrats looking for more discretionary power and larger budgets. In the end, by reducing property rights and suffocating competition in the area of corporate governance one actually harms the interests of the weaker parties.

Garrouste (Chapter 17) also deals with the theory of the firm and contrasts the traditional view, which holds that property rights are instrumental in internalising externalities in production, with the new approach, whereby the firm remains a more or less centralised ownership structure issuing instructions aimed at maximising the owner's target. Decentralised decision making is relegated to areas where uncertainty and incomplete contracts prevail. Such recent theories raise a number of questions about the desirability of voluntary contracts. Nevertheless, although Garrouste credits these new theories with drawing attention to a number of relatively new issues, he believes that at present normative intervention to enhance efficiency is still far from justified.

The remaining chapters of the book focus on intellectual property rights. Isaac and Park (Chapter 18) discuss the foundations of such rights and suggest a free and open system solution (no protection for property). Antonelli (Chapter 19) and Brousseau (Chapter 20) examine the positive and normative economics of knowledge from a neoclassical standpoint, leading to *ad hoc* solutions elaborated by regulatory bodies. Cuccia and Santagata (Chapter 21) provide a case study describing the problems met when designing suitable collective 'cultural' property rights. Finally, Amacher and Meiners (Chapter 22) look at a different sort of intellectual capital – education.

Isaac and Park address the economics of the most important set of property rights to protect inventions – patents. The traditional view is well known. On the one hand, patents encourage technological progress by offering at least temporary monopoly power on the result of innovation (rather than on abstract ideas, which are not patentable). On the other, they often create barriers to entry for other innovators, who are prevented from developing new products starting from somebody else's discoveries. As a result, Isaac and Park caution against property rights to pure or 'near' ideas. In their view pure ideas or near ideas should remain unpatentable, promoting freedom to innovate by making use of other's ideas.

After reviewing the principles and the considerable ambiguities characterising the existing legislation on patents – say, in the software industry – Isaac and Park show how a free and open system enhances the use of the existing technologies, but does not necessarily discourage innovation. This occurs not only when innovators carry out research and development (R&D) with relatively little concern for monetary rewards but also when free and open development proves complimentary to other profitable activities. In other words, according to Isaac and Park, patents promote market power and thus entail static social costs, but they are not guaranteed to provide offsetting dynamic gains. On the contrary, the development of free technology may offer new opportunities for monetary profits to a large number of companies, which may in turn further promote the use of existing know-how and also create new technologies.

Still, Isaac and Park conclude that free and open systems are not inevitably in conflict with proprietary systems, unless innovators following one of the two approaches can prevent others from carrying out their own innovation – which is more than just a theoretical possibility. Indeed, the assessment of the interaction between these two systems remains open and in many respects elusive.

Antonelli's work analyses the economic nature of knowledge in greater depth. In particular, since knowledge is the outcome of a learning process that takes place almost exclusively within the firm, production functions are no longer viewed as well-known blueprints more or less accessible to all managers. Instead, they become highly dependent on the learning path typical of each firm. Hence, they are firm specific, appropriable and tradable. From this standpoint, Antonelli maintains that the market for knowledge is not much different from any other market, and that property rights economics should apply to the production and trade of knowledge, too. In particular, different property rights features as regards their definition, assignment and exclusivity affect the way firms behave: make-or-buy and make-or-sell choices on the one hand, selecting different patterns of integration on the other. In short, knowledge is no public good. Thus, and contrary to what is argued by much of the literature following the Arrow tradition until some two decades ago, subsidies, public procurement and state ownership in key R&D industries can no longer be regarded as the basic instruments to enhance technological progress.

Although Antonelli is aware of the fact that the enforcement of private property rights enhances the production of knowledge, he does not overlook the possibility that such rights might lead to externalities, for by creating monopoly power the diffusion of knowledge is reduced and progress retarded. This is actually the core of the economics of patents, as mentioned in the previous chapters of the *Companion*. By referring the reader to the set of cases known as 'essential facilities' within telecommunication networks, Antonelli concludes that property rights on knowledge – which includes the whole area of intellectual property rights – should thus be weakened by regulation, and (neoclassical) competition enhanced. One example along such lines is compulsory licensing at 'fair prices', to be decided by the judiciary *ex post*, that is, after the new knowledge has been created.

Brousseau also focuses on an area – digital technologies – where marginal costs of production are low and the case for a revival of the economics of public goods seems to be stronger. In particular, Brousseau argues that in this area a fully decentralised property rights system would not be desirable. Hence, his argument for the establishment of a central authority with the responsibility of fixing and monitoring the rules of the game. In other words, a central authority would establish and – if necessary – assign property

rights. Although the author is aware of the costs involved, he believes that in the case of the Internet the advantages of a centralised solution in terms of fairness, better management of addressing systems, economies of scale, economies of scope and cumulative learning processes would prevail.

Another set of situations that are still looking for suitable property rights solutions is analysed by Cuccia and Santagata. They observe what happens when the demand for a precise definition and assignment of intellectual property rights intensifies in a so-called 'cultural district'.[13] After a detailed description of the origins of a typical cultural district (Caltagirone, Sicily), and of the way its production results from specific historical and geographical traits, the chapter explains how a collective trade mark characterising culture-intensive goods may be an effective way to enhance market power, reduce transaction costs and keep out low-quality free riders. However, the authors discover that although producers do acknowledge the need for protection against counterfeiting and free riding, cooperation is viewed with scepticism and the creation of a collective trademark seems not to be particularly welcome. Craftsmen definitely prefer to sign their products individually, piece by piece, and little importance is attributed to the fact that most of the reputation is linked to the district, rather than to the creator. Cuccia and Santagata analyse this behaviour by means of a game-theory exercise, whereby it is shown that a collective trademark is acceptable only if it is designed to convey information about the geographical origin, while remaining clearly distinct from quality certification. But since geographical origin is attributed little importance, all attempts to create and enforce property rights on the cultural district are likely to be in vain.

The final chapter of the volume suggests that fine-tuning collective property rights in the areas of culture and intellectual property may well rank among the great challenges for the years to come. As Amacher and Meiners show, the domain of higher education is a clear example of a situation where transaction costs and perverted incentive structures make it difficult for a residual claimant to emerge. Therefore, making a profit or suffering a loss make little difference. Governing bodies end up by minimising tensions and drastic changes. The solution of controversial problems tends to be postponed, compromises are usually looked for, accepted and justified through 'democratic' procedures. Lack of clear property rights is once again the origin of waste.

Notes

1. Formal rules emanate from a legislative authority through a top-down procedure, whereas informal rules – culture, traditions, customs – are the product of a slow bottom-up process elaborated by a community through repeated interaction.
2. That is, if one believes that ideology shapes moral standards.
3. The so-called 'historical school' and its modern versions argue that economists should be

concerned with little beyond fact collection and more or less sophisticated quantitative analyses, with no ambition to explain complex phenomena. This approach is now by and large discredited. Most would agree that the major concern of the economics profession should be understanding and sometimes influencing the behaviour of individuals in a context characterised by scarce resources.

4. In order to claim perfect knowledge one should have a theory about the end of human scarcity or about the end of humankind's desires to improve their condition, while in order to protect property rights perfectly, one should assume zero transaction costs.
5. This is the first-user–first-owner rule that Hoppe refers to. For instance, it explains why the Spanish Crown's claims to hold property rights over the whole New World as from 1492 were meaningless. The Spaniards were not the first users (other people were there), and were not using most of the territory they were claiming.
6. Of course, systems can also compete, rather than clash. However, Voigt agrees with Harold Berman in saying that this is not very realistic in today's political environments, which seem more prone to centralisation and/or transnational harmonisation, and less favourable to institutional competition.
7. But the author also recalls Berman's work, which showed that situations characterised by competing systems lead to growth, as the history of the Western world has shown.
8. According to the personality theory, property rights are instrumental in allowing the individual to develop and complete his/her personality, for by appropriating a commodity people are able to expand and develop their own traits.
9. Of course, the right to enjoy property (which is morally justified) differs from the desire not to see the value of property fall (which justifies no protection). Theft belongs to the first category; while physical or technological decay belongs to the second.
10. When it comes to germline engineering, Cohen notices that, except in truly exceptional circumstances, everybody is a winner, and the poor probably more so than the rich.
11. Contrary to the functioning of the market process, political action concentrates the benefits and disperses the costs. Although these are usually much larger than the benefits, the benefits become clearly discernible, while the costs tend to go unnoticed.
12. That is equivalent to claiming a property right on the information created within the company, and at the same time to preventing managers or specific shareholders (for example, board members) to make use of such information by disseminating it.
13. A cultural district is defined as an area featuring a unique cultural heritage that translates into entrepreneurship and manufacturing with distinctive elements originating from that local culture. According to Cuccia and Santagata, and contrary to what can be observed in industrial districts, economic development in cultural districts depends heavily on the collective exploitation of the intangible assets derived from the common cultural heritage. Hence the potential need to protect collective property rights.

PART I

THE BIRTH AND EVOLUTION OF PROPERTY RIGHTS

1 Cultural and religious foundations of private property

Leonard P. Liggio and Alejandro A. Chafuen

> Fashion, discipline, and education, have put eminent differences in the ages of several countries, and made one generation much differ from another in arts and sciences. But truth is always the same; time alters it not, nor is it better or worse for being of ancient or modern tradition. (John Locke [1697] 1802, Section 24, p. 79)

Introduction

Private property is both a notion and a reality, which has concrete existence. In many different cultures, and from many different religious perspectives, people have analysed, justified and recommended different degrees of respect for appropriation. The economic arguments in favour of private property are important. Most economists agree that economics is a value-free science. Valuing property falls outside economic science. In order for private property to become a reality it will need more: it will need to be embodied in a rule of law, a culture that respects it, or at least leaders who value and are ready to implement it.

Although we acknowledge that actions are at least as relevant as words to understand the institution of private property, in this chapter we shall focus on the role of ideas. This is an essay on the history of ideas placed in a historical context of Western civilization.

While economists might have credibility when arguing from a cost–benefit perspective, moralists have a much higher degree of authority when arguing about values. In recent surveys, religious leaders appear as the most respected figures in many societies. This is true in countries with large Muslim populations as well as in many Latin American ones. That was also true in England and in the United States during the period when private property evolved as an essential institution.

Economics is not the only factor influencing cultures. As John Locke reminded us:

> [O]ur Savior found mankind under a corruption of manners and principles, which ages after ages had prevailed, and must be confessed, was not in a way or tendency to be mended. The rules of morality were in different countries and sects different. And natural reason nowhere had cured, nor was like to cure, the defects and errors in them. Those just measures of right and wrong, which necessity had

> anywhere introduced, the civil laws prescribed, or philosophy recommended, stood on their true foundations. They were looked on as bonds of society, and conveniences of common life, and laudable practices. But where was it that their obligation was thoroughly known and allowed, and they received as precepts of law – of the highest law, the law of nature? That could not be, without a clear knowledge and acknowledgement of the lawmaker, and the great rewards and punishments for those that would or would not obey him. But the religion of the heathens, as was before observed, little concerned itself in their morals. The priests, that delivered the oracles of heaven, and pretended to speak from the gods, spoke little of virtue and a good life. And, on the other side, the philosophers, who spoke from reason, made not much mention of the Deity in their ethics. They depended on reason and her oracles, which contain nothing but truth, but yet some parts of that truth lie too deep for our natural powers easily to reach and make plain and visible to mankind without some light from above to direct them. (Locke [1695] 1989, pp. 176–7)

Economists also rely on reason for their oracles. Some are aware of their limitations. Peter Bauer and Douglass North recognized that some of the essential elements needed for development are factors that are beyond economics. Honesty, trust, integrity and common decency are needed for the emergence, dissemination and effective implementation of the institutions of private property. Yet economists have no special training to understand these traits of human behaviour.

John Locke, who spent the last years of his productive life to better understand the Holy Scriptures and explaining the benefits of Christianity, was also a great champion of theology. This was a science that stood 'incomparably above all the rest', which has as its scope 'the honour and veneration of the Creator, and the happiness of mankind. This is that noble study which is every man's duty, and every one that can be called a rational creature is capable of' (Locke [1697] 1802, p. 72). Locke added: 'This is that science which would truly enlarge men's minds, were it studied, or permitted to be studied everywhere with that freedom, love of truth, and charity which it teaches, and were not made, contrary to its nature, the occasion of strife, faction, malignity, and narrow impositions'(ibid., p. 73).

He also lived his faith well:

> As he was incapable for a considerable time of going to church, he thought proper to receive the sacrament at home, and two of his friends communicated with him; as soon as the office was finished, he told the minister that he was in the sentiments of perfect charity towards all men, and of a sincere union with the Church of Christ, under whatever name distinguished. (ibid., p. vii)

The same biographer states that he spent his time in 'acts of piety and devotion' exhorting those at his bed-side that this life should only be regarded as a preparation for a better (ibid., p. vii).[1]

Religious writers and, more important, people who lived religious principles promoting honesty, integrity and common decency, have all contributed to the emergence of private property. The maxim *res non verba* is more important in issues of culture and religion than in any other field. The kingdom of heaven is reserved for those who practise the right principles, not to those who claim 'Lord, Lord'. J.G. Courcelle-Seneuil (1813–92) wrote that ethics is the essential part of religion, because it is its practical side, and more relevant for worldly discussions. He also wrote that citizens help mould ruling opinions by their spoken and written words but, most of all, by their actions (Courcelle-Seneuil 1887, p. 325). Antoine Yves Goguet (1716–58) was one of the most influential European authors on the origin of laws, writing after Locke, his work, *The Origins of Laws, Arts, and Sciences, and their Progress Among the Most Ancient Nations* (Goguet [1758] 1775) was used extensively in the first edition of the *Encyclopaedia Britannica*, and the translation of his book was done in Edinburgh prior to Adam Smith's publication of *The Wealth of Nations*. One English edition appeared in 1761 and another in 1775. At the time, translations also appeared in German, Spanish and Italian.

According to Goguet,

> We are not therefore to consider the first laws of society, as the fruit of any deliberation, confirmed by solemn and premeditated acts. They were naturally established by a tacit consent, a kind of engagement to which men are naturally very much inclined. Even political authority was established in this manner, by a tacit agreement between those who submitted to it, and those who exercised it. This kind of tacit agreement was also the origin of those CUSTOMS, which, for a long time, were the only laws known among mankind. (Goguet [1758] 1775, vol. I, p. 8)

> There was no need of any particular solemnities in establishing such rules and maxims as these. They derive their origin from those sentiments of equity and justice which GOD has ingraved [sic] in the hearts of all men. They are taught us by that internal light, which enables us to distinguish between right and wrong: dictated by that voice of nature, which will make itself be heard, or will alarm the soul with tormenting remorse as often as its dictates are disobeyed. (ibid., vol. I, p. 7)

Goguet credits the laws of marriage as the beginning of civilized life: 'These laws, in a word, by ascertaining the rights of children, have secured a succession of subjects to the state, and given a regular and settled form to society. No kind of laws have contributed more than these to preserve peace and harmony amongst mankind'(ibid., vol. I, p. 72). 'The institution of the rights of property, and the laws of marriage, necessarily introduced certain customs and usages which may be regarded as the foundation and origin of all civil laws'(ibid., vol. I. p. 25).

Goguet then describes many societies that relied only on customs, especially in the Indies, yet that was not enough to encourage development. He also argued that laws protecting property and setting boundaries were essential for society. He studied Father José de Acosta's history of the Indies, and relied on his reports to state that even the first Incas in Peru 'took great pains, in dividing and distributing the lands amongst their subjects' (ibid., vol. I, pp. 30–31).

He adds 'It was not enough to establish and regulate the division of lands, it was also necessary to suppress and prevent usurpations'. He continued showing all the biblical passages calling for security of possessions and clear boundaries. Agriculture again reinforced the need for private property: 'the toil and labour which the cultivation of land requires, gave men a strong attachment to what cost them so much fatigue'. Giving the results of their work to their children required rights of inheritance 'thus the division of lands gave rise to rights and to jurisprudence' (ibid., vol. I, pp. 31–2).

Property in antiquity: marriage, family and law

The great English legalist in the age of the American revolution, Sir William Blackstone (1723–80) defined property as 'that sole and despotic dominion which one man claims and exercises over the external things of the world, in total exclusion of the right of any other individual in the universe'(Blackstone [1765–69] 1962, II, I, p. 2).

Property is an institution of the human person from the first record of his customs. 'The law finds the institution of property in existence, as well at the earliest as at all later stages of growth, and, far from creating its varieties, is occupied only in defining, maintaining and validating them' (Noyes 1936, p. 18).

Sir Henry Sumner Maine (1822–88), in his *Ancient Law* ([1861] 1986) described the common legal customs of the Indo-European peoples, drawing on his judicial experience in India. The Indo-European legal history that has received the greatest study is that of the ancient Greeks and Italians.

History professor at the Sorbonne, Numa Denis Fustel de Coulanges (1830–89), *The Ancient City: A Study of the Religion, Laws, and Institutions of Greece and Rome* (1864), provides an analysis of the Classical legal institutions. The Indo-Europeans did not believe that after life there was no afterlife; rather, they looked on death not as a termination, but as a change of life to a second existence. In a religious tradition in which the souls of the dead remained near the living families, the graves of the dead members of the family and their daily remembrance by the head of the family were of central importance.

Each Greek and Italian family possessed its own religion because each had their own particular ancestors to whom daily respect was offered. The family

altar and fire were the focus of religion of each family. Observance of the daily rites was necessary for the happiness of the dead ancestors and thus of the success of the living family. The family home and land were an extension of the family ancestors. 'The members of the ancient family were united by something more powerful than birth, affection, or physical strength; this was the religion of the sacred fire, and of dead ancestors. This caused the family to form a single body, both in this life and in the next' (Fustel de Coulanges [1864] 1956, p. 42).

Marriage was the first institution of each domestic religion. The wife became a priestess of her husband's domestic religion and, should she die, the widower could no longer perform his priestly functions. The worship of the ancestors and the domestic fire was transmitted from male to male, but was shared by the wife. By the sacred ceremony of marriage,

> [The husband] is now about to bring a stranger to this hearth; with her he will perform the mysterious ceremonies of his worship; he will reveal the rites and formulas which are the patrimony of his family. There is nothing more precious than this heritage; these gods, these rites, these hymns which he has received from his fathers, are what protects him in this life, and promise him riches, happiness, and virtue. And yet, instead of keeping to himself this tutelary power, as the savage keeps his idol or his amulet, he is going to admit a woman to share it with him. (ibid., p. 43)

By the sacred ceremony of marriage the wife is ordained and adopted into the domestic religion as a necessity for her to become a priestess of the sacred fire of her husband's ancestors. The marriage ceremony at the sacred fire culminates in the husband and wife sharing a wheaten loaf:

> This sort of light meal, which commences and ends with a libation and a prayer; this sharing of nourishment in presence of the fire; puts the husband and wife in religious communion with each other, and in communion with the domestic gods'(ibid., p. 46)

> The institution of sacred marriage must be as old in the Indo-European race as the domestic religion; for the one could not exist without the other. This religion taught man that the conjugal union was something more than a relation of the sexes and a fleeting affection, and united man and wife by the powerful bond of the same worship and the same belief. The marriage ceremony, too, was so solemn, and produced effects so grave, that it is not surprising that these men did not think it permitted or possible to have more than one wife in each house. Such a religion could not admit of polygamy. (ibid., pp. 47–8)

Fustel de Coulanges notes:

> There are three things which, from the most ancient times, we find founded and solidly established in these Greek and Italian societies: the domestic religion; the

> family; and the right of property – three things which had in the beginning a manifest relation, and which appear to have been inseparable. The idea of private property existed in the religion itself.
>
> Every family had its hearth and its ancestors. These gods could be adored only by this family, and protected it alone. They were its property. Now, between these gods and the soil, men of the early ages saw a mysterious relation. Let us first take the hearth. This altar is the symbol of a sedentary life; its name indicates this. It must be placed upon the ground; once established, it cannot be moved. The god of the family wishes to have a fixed abode … (ibid., p. 61)

He shows the role of religion in property:

> It did not matter whether this enclosure was a hedge, a wall of wood, or one of stone. Whatever it was, it marked the limit, which separated the domain of one sacred fire from that of another. This enclosure was deemed sacred. It was an impious act to pass it. The god watched over it, and kept it under his care. … This enclosure, traced and protected by religion, was the most certain emblem, the most undoubted mark of the right of property.
>
> Let us return to the primitive ages of the Aryan race. The sacred enclosure, which the Greeks call *eoxos*, and the Latins *herctum*, was the somewhat spacious enclosure in which the family had its house, its flocks, and the small field that it cultivated. In the midst rose the protecting fire-god. Let us descend to the succeeding ages. The tribes have reached Greece and Italy, and have built cities. The dwellings are brought nearer together; they are not, however, contiguous. The sacred enclosure still exists, but it is of smaller proportions; oftenest it is reduced to a low wall, a ditch, a furrow, or to a mere open space, a few feet wide. But in no case could two houses be joined to each other; a party wall was supposed to be an impossible thing. The same wall could not be common to two houses; for then the sacred enclosure of the gods would have disappeared. At Rome the law fixed two feet and a half as the width of the free space, which was always to separate two houses, and this space was consecrated to 'the god of the enclosure'. (ibid., pp. 62–3)

Each family home is a domestic temple, which gives a sacred character to the land, which surrounds and encompasses it. The family is consecrated master and proprietor of the land of the domestic divinities. The right of private property is a sacred right.

'What is there more holy', says Cicero, 'what is there more carefully fenced round every description of religious respect, than the house of each individual citizen? Here is his altar, here is his hearth, here are his household gods; here all his sacred rights, all his religious ceremonies, are preserved' (Cicero, *Pro Domo*, in Fustel de Coulanges [1864] 1956, p. 64).

The boundary of each property was marked by an upright post or stone, a *terminus* which was considered divine as part of the family religion:

> The employment of *Termini*, or sacred bounds for fields, appears to have been universal among the Indo-European race. It existed among the Hindus at a very early date, and the sacred ceremonies of the boundaries had among them a great analogy with those which Sculus Flaccus has described for Italy. Before the foundation of Rome, we find the *Terminus* among the Sabines; we also find it among the Etruscans. The Hellenes, too, had sacred landmarks … (Fustel de Coulanges [1864] 1956, p. 68)

> To encroach upon the field of a family, it was necessary to overturn or displace a boundary mark, and this boundary mark was a god. The sacrilege was horrible, and the chastisement severe. According to the old Roman law, the man and the oxen who touched a *Terminus* was devoted – that is to say, both man and oxen were immolated in expiation. The Etruscan law, speaking in the name of religion, says, 'He who shall have touched or displaced a bound shall be condemned by the gods; his house shall disappear; his race shall be extinguished; his land shall no longer produce fruits; hail, rust, and the fires of the dog-star shall destroy his harvests; the limbs of the guilty one shall become covered with ulcers, and shall waste away.' (ibid., p. 69)

> Plato, *Laws*, VIII, p. 842 states: 'Our first law ought to be this: let no person touch the bounds which separate his field from that of his neighbor, for this ought to remain immovable. … Let no one attempt to disturb the small stone which separates friendship from enmity, and which the land-owners have bound themselves by an oath to leave in its place'. (ibid., p. 69)

The Gospels

In the Gospels, Jesus makes various references to wealth. Some of the parables speak of the value of investment in property, trade and human capital as examples of spiritual investment. Many of the recommendations of Jesus are aimed at those who wish to join his circle of disciples to live a rigorous life. Often they are counsels of perfection, and not aimed at ordinary believers who live their everyday life in their family, their work and their prayers. These counsels of perfection are calls to a special vocation of a spiritual life; it is the counsels of perfection – chastity, poverty and obedience – which are followed by the members of religious orders – monks, friars, canons regular[2] brothers and nuns – in the Orthodox, Catholic, Anglican and Lutheran churches.

In particular, Jesus' counsels of perfection can be found in several of the Gospels. In Luke 18:1–8, Jesus recommends persistent prayer to God. In Luke 18:9–14, Jesus recommends the quiet prayer of the sinner (tax collector) standing far off in the Temple as opposed to the self-congratulatory prayer of the Pharisee. Christian tradition interpreted Luke's next verses (18:15–17) to enjoin the perfection of obedience like little children (also, Mark 10:13–16 and Matthew 19:13–15). Similarly, in the second counsel of perfection regarding celibacy in Mark 10:1–12 and Matthew 19:1–2 Jesus declares: 'All receive not this word but they whom it is given. … He that can receive it, let him receive it'.

Then the matter of poverty for the perfect is addressed in Luke 18:18–30, Mark 10:17–31 and Matthew 19:16–30. The rich man seeks Jesus' counsel on going beyond the normal commandments. Jesus declares:

> If thou wilt be perfect, go sell what thou hast, and give to the poor; and thou shalt have treasure in heaven; and come, follow Me. … And Jesus seeing him become sorrowful, looking round about, said to His disciples: Amen, I say to you: How hard shall they who have riches enter into the kingdom of God. And the disciples were astonished at His words. But Jesus answering again, said to them: Children, how hard it is for them that trust in riches to enter the kingdom of God. Again I say to you: It is easier for a camel to pass through the eye of a needle than for a rich man to enter into the kingdom of heaven. And when the disciples heard this, they wondered the more, saying among themselves: Who, then, can be saved? And Jesus looking on them, said: With men it is impossible, but not with God; for all things are possible with God.

A minority interpretation among scripture scholars notes that 'the eye of the needle' was the name given to the pedestrian gate which would require the camel to pass through on its knees; something that would require extra effort.

Jewish and Christian thought on the market and money in Hellenistic and Roman ages

If one examines the Old Testament, one finds references to issues such as private property. In the earliest books of the Old Testament, when the Hebrews arrived in the Holy Land, land was distributed to them as individual holdings, and they were enjoined under penalty of sin from moving boundaries from the land or changing the boundaries. That would amount to theft as in coveting a neighbour's goods. In the Old Testament there is an emphasis that one should not make an idol of property, just as one should not make an idol of poverty; that material goods should not come before one's obligations to God.

In the Old Testament there are references regarding the unjust taking of property. In Isaiah 1:23 the prophet warns princes not to consort with thieves or give corrupt judgments. This is preceded, in Isaiah 1:22, by the statement: 'Thy silver has become dross; thy wine has become mixed with water'. Isaiah is warning here against the dilution of currency, and we see that in many other places in the Old Testament, the Prophets are condemning this dilution in the rules (in essence, inflation), and treating it as a major form of theft alongside the princes' taking-away of private property. Ezekiel 22:18–22 used the evil of debasing coin as an example for princes to address in their individual reformation.

The Old Testament prophets placed great emphasis on defending the individual family's right of property against the state. Ezekiel warns against the oppression whereby the property owned by the individual Jew is taken by the

ruler. This is a major theme that then continues in the early Christian literature.

During the Hellenistic and Roman periods there was a division in Judea between the Sadducees (who were from the priestly caste and mainly lived in Jerusalem) and the Pharisees and the other pious people in the countryside. The Sadducees oversaw the collection of taxes, and were the object of the criticism directed at the rich (tax collectors). The Pharisees were developing a belief in immortality of the soul and bodily resurrection, and a strong sense of the role of oral tradition in addition to the written Bible. When one comes to the birth of Jesus and his emergence in public life, Jesus is articulating the language of the Pharisees, while not their legalistic practices. Jesus' public ministry was supported by the wealth of his friends and disciples. Wealth was not condemned as it was necessary to support Jesus' public ministry; people were not condemned by failure to use their wealth to support Jesus and his disciples. This use of wealth was a special calling.

In the Epistles of Paul, we find a continuity of Stoic ideas, some of which are very similar to Christian ideas. In particular, Paul refers to the Stoic idea of the importance of self-sufficiency – the importance of people having their property in order to be self-sufficient and working to achieve enough property to support their family. If Christians do not work to achieve that necessary wealth they are lacking in the necessary Christian grace.

A leading Father of the early Church was the Athenian-born Clement of Alexandria (AD150–215). Alexandria was the great city of the Eastern Mediterranean. Founded by one of Alexander the Great's generals, Ptolemy I Soter (367–283BC), Alexandria became the centre of Greek philosophy associated with its great library. Ptolemy II Philadelphus (309–246BC) asked 70 scholars of the city's large Hebrew community to translate the books of the Bible into Greek. They created the *Septuagent* Greek Bible (270BC).

Titus Flavius Clement of Alexandria was head of the Christian school in which Origen was one of his pupils. Clement left Alexandria during the persecution of Emperor Septimus Severus. He approved private property and the accumulation of property in his *Who Is the Rich Man that is Saved?* in which he analyses Mark 10:17–31. He did not encourage ordinary Christians to pursue an ascetic ideal of giving up one's possessions. Clement argued that riches and goods are the means that can benefit our neighbour. They have been provided by God for the good of humankind to be used by those who know how to use them.[3]

Clement was not impressed with the argument that poverty equalled holiness. He noted that if poverty equalled holiness, then proletarians, derelicts and beggars, and some who have few virtues and are ignorant about God, would be the best candidates for religious life simply because they had no money.[4]

The Christian Father Lactantius (AD260–340) emphasized that the concept of abolition of private property was unacceptable to Christians. He noted that it was Plato who introduced the notion of community of property rooted in the unnatural idea of community of wives. As a Christian Father, Lactantius calls out to the Christian world that it must first protect private property if it wants to protect the family from all the assaults of the state. Lactantius joined other philosophers in considering private property as a distinguishing quality of humankind, as opposed to animals. Lactantius explained the interconnection between private ownership and the virtues that come from it, the sound families it produces, and how the assault on private property is an assault on Christian virtue.

Wealth exists, according to the Old Testament, to be used productively and wisely, and this theme is continued in the New Testament. In I Timothy 1:3, there is a strong statement about the responsibility of the family to produce wealth and thereby to care for its own. In order to defend the family, it is imperative to defend private property. Lactantius stated: 'for ownership of things contains the matter of virtues and of vices, but community holds nothing other than the license of vices' (Lactantius 1871).[5]

St Augustine of Hippo (AD354–430) responded to his close friend from Syracuse, Hilarius, regarding the remarks of some Christians in Syracuse: 'That a rich man who continues to live richly cannot enter the Kingdom of Heaven unless he sells all he has, and that it cannot do him any good to keep the Commandments while keeping his riches'.

Augustine responded at great length in order to supply Hilarius with arguments to rebut this idea. Augustine declared:

> Listen, now, to something about riches in answer to the inquiry in your next letter. In it you wrote that some are saying that a rich man who continues to live richly cannot enter the Kingdom of Heaven unless he sells all that he has, and that it cannot do him any good to keep the Commandments while he keeps his riches. Their arguments have overlooked our Fathers, Abraham, Isaac, and Jacob, who departed long ago from this life. It is a fact that all these had extensive riches, as the Scripture faithfully bears witness, yet He who became poor for our sake, although He was truly rich, foretold in a manner in a truthful promise that many would come from the East and West and would sit down, not above them, not without them, but with them in the Kingdom of Heaven. Although the haughty rich man who is clothed in purple and fine linen, and feasted sumptuously every day, died and was tormented in Hell, nevertheless, if he had shown mercy to the poor man covered with sores who lay at his door, and was treated with scorn, he himself would have deserved mercy. And if the poor man's merit had been his poverty, not his goodness, he would surely not have been carried by Angels into the Bosom of Abraham, who had been rich in this life. (Augustine [fifth century] 1951, p. 340)[6]

Augustine continues his analysis of the example of our Father Abraham, and then supposes that the Christians of Syracuse probably say that the patriarchs

did not sell all they had to give it to the poor because the Lord had not commanded it. Augustine says:

> We believe that the Apostle Paul was the minister of the New Testament when he wrote to Timothy, saying, 'Charge the rich of this world not to be high-minded, nor to trust in the uncertainty of riches, but in the Living God, who giveth us abundantly all things to enjoy. To do good, to be rich in good works, to give easily to communicate to others. To lay up on store for themselves a good foundation against the time to come, that they may hold on to the true life' (1 Tim. 6.17–19), in the same way as it was said to the young man, 'If thou will enter into life.' I think when he gave those instructions to the rich, the apostle was not wrong in not saying, 'Charge the rich of this world to sell all they have and give to the poor and follow the Lord', instead of, 'Not to be high-minded nor to trust in the uncertainty of riches.' It was his pride, and not the riches, that brought the rich young man to the torment of Hell because he despised the good poor man who lay at his gate; because he put his hope in the uncertainty of riches and thought himself happy in his purple and fine linen and sumptuous banquet. (ibid., pp. 342–3)

Augustine said that it was unlawful to steal to give alms. The medieval *Decretals* imposed a penance of three weeks upon a man who commits theft because he is hungry.

The Christian Father, Salvian's (AD405–95) major work, *De Gubernatione Dei* (The government of God) was completed in about AD450 in Marseilles. Salvian asks: 'What is a political position but a kind of plunder? There is no greater pillage of poor states than that done by those in power'. He continues to speak of tax collectors by speaking of those strangled by the chains of taxation as if by the hands of brigands. 'There is found a great number of the rich whose taxes kill the poor' (Salvian 1947, pp. 100–101).[7] Salvian continued:

> What towns, as well as municipalities and villages, are there in which there are not as many tyrants as tax collectors? Perhaps they glory in the name of tyrant because it seems to be considered powerful and honored. … What place is there, as I have said, where the bowels of widows and orphans are not devoured by the leading men of the city, and with them almost all Holy Men? … They seek among the barbarians the dignity of the Roman because they cannot bear barbarous indignity among the Romans. (ibid., pp. 134–5)

The Roman emperors appointed the rich as tax collectors who were responsible from their own wealth for the annual tax burden. The system was built around the collection and the avoidance of taxes. To escape the tax burdens, and the tortures associated with tax collection, many people fled from their farms and from the cities, and lived in the countryside or in the wilderness. Many of these were very religious people, and some of the early monastic communities, as in Egypt, evolved from these refugees from taxation. Many rich and poor fled taxation by moving to the areas ruled by the Germanic tribes which had migrated into the Roman provinces.

Salvian declared:

> Thus, far and wide, they migrate either to the Goths or to the Baghudi, or to other barbarians everywhere in power, yet they do not repent having migrated. They prefer to live as free men under an outward form of captivity, than as captives under appearance of liberty. Therefore, the name Roman citizens, not only greatly valued but dearly bought, is now repudiated and fled from – and it is considered not only base but ever deserving of abhorrence. And what cannot be a greater testimony of Roman wickedness than that many men, upright and noble and to whom the position of being a Roman citizen should be considered as of the highest splendor and dignity, have been driven by the cruelty of Roman wickedness to such a state of mind that they do not wish to be Romans? … [t]hey, who suffer the insistent, and even continuous destruction of public tax levies, to them there is always imminent a heavy and relentless proscription. They desert their homes, lest they be tortured in their very homes. They seek exile lest they suffer torture. The enemy is more lenient to them than the tax-collectors. This proved by this very fact that they flee to the enemy in order to avoid the full force of the heavy tax levy. (ibid., pp. 136, 138)

Germanic law codes in Western Europe

When the Germanic tribes sought protection from the Huns by entering the Roman Empire, they settled in various provinces where they were allotted lands as formal guests of the Roman landowners. One major difference between the German tribes and the Romans was that the Germanic tribes (except the Franks) had already been converted to Christianity by Arian missionaries, before their migrations. The laws of the various German tribes were written as law codes (*leges barbarorum*), which applied to cases between Germans, or between Germans and Romans, while cases between Romans were decided according to the Roman laws (*leges romanae*) in existence in the Western Roman Empire preceding the deposition of the last Western Roman Emperor in AD476. The Eastern Roman Empire or Byzantine Empire in Constantinople continued until AD1453 and developed its separate legal code (AD528–34) under the Emperor Justinian I (who ruled from AD527–65). One of the earliest Germanic codes was the *Burgundian Code*, applied to the Rhone region of the Roman province of *Gallia Narbonensis*. The earlier Roman *Codices* were compiled in the *Lex Romana Burgundionum*. For the Burgundians, the code was the *Lex Burgundionum*, also called the *Lex Legum Gundobaldi* after King Gundobad (who reigned from AD474–516). The *Burgundian Code* declared (IV, 3):

> And if any natural freeman, either Burgundian or Roman, takes in theft a pig, a sheep, a beehive, or a she-goat, let him pay three-fold according as their value is established, and in addition let him pay a fine of twelve *solidi*. Let the composition be for the pig, one *solidus*; for the sheep, one *solidus*; for the beehive, one *solidus*; for the goat, a *tremissis*. Indeed, let their value be paid threefold. (*Burgundian Code* 1949, pp. 24–5, XXVII, 1)[8]

> If a native freeman breaks and opens another's fence when subject to no impedient [impairment] therefrom, only of the purpose of causing such damage, let him pay a single *tremissis* for each stake to him whom the crop belongs. … (ibid., p. 42, LV, 2)

As often as cases arise between two Romans concerning the boundaries of fields which are possessed by barbarians though the law of hospitality, let the guests of the contestants not be involved in the quarrel, but let them await the outcome between the Romans contending in judgment. And the guest of the victor shall have a share of the property obtained as a result of his success. (ibid., pp. 64–5)[9]

The Middle Ages

The institutions of Western Europe during the Middle Ages recognized private property. The strengthened protection of the property of peasants, merchants, religious orders and Jews resulting from the Peace of God and the Truce of God movements in the tenth century permitted a flourishing of agricultural production and widening commerce.[10]

The right of private property in medieval Europe is a backdrop to the discussion of property among the Roman lawyers, canonists and the theologians. The most important new work in the field is by the Cambridge University Social Anthropologist Alan Macfarlane in his *The Origins of English Individualism* (1978). Macfarlane explains how in the twelfth century the kinship and feudal limitations on property began to dissolve. Property began to be saleable on the market for agricultural lands. Among the changes in customs was the impact of Christianity. Max Weber saw Christianity challenging the traditional clan system. According to Weber,

> Christianity, which encouraged an abstract, non-familistic attitude stressed the individual believer; 'every Christian community was basically a confessional association of individual believers, not a ritual association of kinship groups.' This 'all-important destruction of the extended family by the Christian communities …' was the foundation upon which an autonomous bourgeoisie developed in the cities of western Europe. (Macfarlane [1978] 1979, p. 50)[11]

Macfarlane quotes the leading medieval historian, M.M. Postan, on the importance of cash in the rural medieval economy and the money market in land:

> Historians are now agreed that commutation of labour services was by no means a new phenomenon in the late fourteenth century. There was widespread, and on some estates, wholesale, commutation of labour services in the middle of the twelfth century … Professor Kosminsky has recently reminded us that by 1279 – the date of the Hundred Rolls – labour dues no longer were the main source of the lord's income from his peasant tenants. (ibid., p. 151)

Macfarlane concludes: 'so we could describe thirteenth-century England as a capitalist-market economy without factories' (ibid., p. 196). He continues:

> It has been argued that if we use the criteria suggested by Marx, Weber and most economic historians, England was as 'capitalist' in 1250 as it was in 1550 or 1750. That is to say, there were already a developed market and mobility of labour, land was treated as a commodity and full private ownership was established, there was very considerable geographical and social mobility, a complete distinction between farm and family existed, and rational accounting and profit motive were widespread. (ibid., pp. 195–6).
>
> On this topic, see also Randall Collins 1986.[12]

> Yet, if the present thesis is correct, individualism in economic and social life is much older than this in England. In fact, within the recorded period covered by our documents, it is not possible to find a time when an Englishman did not stand alone. Symbolized and shaped by his ego-centred kinship system, he stood in the centre of his world. This means that it is no longer possible to 'explain' the origins of English individualism in terms of either Protestantism, population change, the development of a market economy at the end of the middle ages, or the other factors suggested by the writers cited. Individualism, however defined, predates sixteenth-century changes and can be said to shape them all. The explanation must lie elsewhere, but will remain obscure until we trace the origins even further than has been attempted in this work. (ibid., pp. 196–7)

While the professors in the law schools, such as Bologna, lectured on the civil and canon laws which were based on the right to property of the Roman law, philosophy faculties, such as Paris, where law was not taught, were not anchored in the reality of Roman property law. The philosophers of the eleventh and twelfth centuries were mainly members of religious orders, which meant that they had taken vows of poverty, celibacy and obedience. These they viewed as the best way to achieve Heaven, and they tried to convince the ordinary Christian to adopt the same route. Thus, they wished to place the private ownership of property in a negative light from a religious point of view, and to condemn it. Their arguments did not seek the virtue of consistency.

Some argued: whatever belonged to no one should belong to the person who occupies it or takes possession of it. They said it was necessary that private property exist, because if all were goods were held in common, the wicked would take everything and the good would suffer deprivation. Private property was the natural protection of the life and rights of the good and the whole of humankind against the corruption of the wicked.

Pope Innocent IV, born Sinibaldo de' Fieschi, in Genoa, Italy, taught at the University of Bologna and was created a cardinal in 1227 and elected pope in 1243; he died in 1254. In his *Apparatus ad Quinque Libros Decretalium* (edition Venice, 1578) or *Commentary on the Decretals* (III, 34, 8), he said

that private property arose from the custom of our first ancestors. He declared that private property was good, and was not evil, because goods which were held in common were neglected by all, and goods held in common tended to lead to discord. Private property was a source of harmony among humankind. Canonists saw private property as not part of the primeval law of the animal world, but of the natural law of man created in the image and likeness of God. Thus, it was natural that our first ancestors exercised the right of private property.

In contrast to earlier monastic writers, who saw wealth as equal to sin and poverty as equal to holiness, St Thomas Aquinas (born Roccasecca, 1225, died Fossa Nuova, 1274) introduced the rationalism of Aristotle's writings recently translated from Arabic sources. In the scholastic method, Aquinas presents all the arguments on each side of a question. He then accepts the previous authorities and at the same time introduces new arguments which move the substance to a new level of sophistication away from the monastic pastoralisms.

Aquinas, a Dominican Friar, recognized individual self-ownership, a 'proprietary right over himself', as founded on man being a rational being. He grounded the right of original acquisition on labour and occupation. The cultivation of unclaimed land created a just property title in that land in that one man. A further statement of Aquinas' argument was made by his leading student, John of Paris (Jean Quidort, c. 1250–1306) who explained the absolute right of private property:

> [Property is] acquired by individual people through their own skill, labour and diligence, and individuals, as individuals, have right and power over it and valid lordship; each person may order his own and dispose, administer, hold or alienate it as he wishes, so long as he causes no injury to anyone else; since he is lord. (John of Paris [c. 1302] 1974, p. 28)

He continued:

> Exterior possessions of the laity are not conferred on the community as are ecclesiastical possessions; rather they are acquired by individuals through their own art, labor and industry. Individual persons, in so far as they are individual, have in themselves right, power and true ownership; and each one may deal with his own – by disposing of it, distributing it, retaining it, or alienating it – as he pleases without injuring anyone else, since he is its owner. (ibid., p. 28)

In the *Summa Theologica* (1265–73), Aquinas wrote that private property is necessary for human life for three reasons:

> First, because each person takes more trouble to care for something that is his sole responsibility than what is held in common or by many for in such a case each

> individual shirks the work and leaves the responsibility to somebody else, which is what happens when too many officials are involved. Second, because human affairs are more efficiently organized if each person has his own responsibility to discharge; there would be chaos if everybody cared for everything. Third, because men live together in greater peace where everyone is content with his task [sic] [*re sua contentus est*]. We do, in fact, notice that quarrels often break out amongst men who hold things in common without distinction. [*Inter eos qui communiter et ex indiviso aliquid possident frequentius jurgia oriuntur*]. (cited in Roberts and Donaldson (eds) 1871, vol. 7, pp. 92–3)

It is important to note that in this edition the phrase '*re sua contentus est*' is translated as 'everyone is content with his task'. A more precise version would read 'everyone is content with his things'. In this passage Aquinas stated that satisfaction may be derived from ownership of private property (*Summa Theologica*, II–II, qu. 66, art. 2, resp.)[13]

Aquinas argues that private property is lawful because it is efficient in the use and production of the material things put on earth for the use and pleasure of men, and private property is necessary for the tranquillity of humankind. From the perspective of a Dominican Friar, Aquinas felt that there was much superfluous wealth and advocated that the owners of private property should redistribute the superfluous wealth for the benefit of the destitute.

The Aquinas–John of Paris–John Locke labour theory of private property is completely different from the Marxist labour theory of the value or price of goods. The Aquinas analysis was challenged by the Spiritual Franciscans who reacted against Thomist rationalism and sought to restore fidelism and the practices of some Christians in Jerusalem after the death of Jesus. Franciscan John Duns Scotus (1265–1308) preached that in the state of innocence all goods were held in common and no private property should exist.

These Spiritual Franciscans were condemned by Pope John XXII in the bull, *Quia vir reprobus* (1329) which declared that since man was created in the image and likeness of God, each man's private property reflected God's dominion over his material possessions. Property rights were founded not on positive law or convention, but in man's nature and the natural law. An excellent treatment of the debates over property and poverty concerning the Spiritual Franciscans, as well as of Scholastic analyses of economic issues can be found in Murray Newton Rothbard, *Economic Thought Before Adam Smith* (Rothbard 1995, vol. I, pp. 31–64).

Richard Tuck (1979, p. 27) and Rothbard (1995, pp. 93–5) analyse the important contribution to natural property rights theory made by a chancellor of the University of Paris, Jean Gerson, defender of the university against kings and popes. Gerson (1962) stated in *De Vita Spirituali Animae* (1402):

> There is a natural dominion as a gift from God, by which every creature has a *ius* (right) directly from God to take inferior things into his own use for its own preservation. Each has this *ius* as a result of a fair and irrevocable justice, maintained in its original purity, or a natural integrity. In this way Adam had dominion over the fowls of the air and the fish in the sea … To this dominion the dominion of liberty can also be assimilated, which is an unrestrained faculty given by God.[14]

Aquinas and his successors provided the intellectual tools for philosophers to apply the reason of Aristotle to the increasingly complex issues of the modern world. These achieved their fruition in the writings of the Scholastic philosophers of the Late Medieval period and the Renaissance who provided the foundation for Hugo Grotius, Samuel Pufendorf and John Locke.[15]

The contributions of Late Medieval scholars

Medieval Scholasticism encompassed some seven centuries, from AD800 to AD1500. In theological and philosophical studies the activity of the period from AD1350 to AD1500 is known as Late Scholasticism. In social sciences, Late Scholasticism extends to the end of the seventeenth century.

The Medieval Schoolmen or, as they preferred to be called, the 'Doctors', were the foremost thinkers of their times. Their analyses and conclusions shaped Catholic thinking so persuasively that they continue to be a significant foundation of contemporary Church doctrine.

St Thomas Aquinas was the foremost Scholastic writer. His influence was so widespread that nearly all subsequent Schoolmen studied, quoted and commented upon his remarks. The century following Aquinas produced many Scholastic authors whose works analysed private property. St Bernardino of Siena (1380–1444), St Antonino of Florence (1389–1459), Jean Gerson (1363–1429), Conradus Summenhart (1465–1511) and Sylvestre de Priero (d. 1523) are the best known.

The Hispanic Scholastics

Francisco de Vitoria (c. 1480–1546) is called the Father of the Hispanic Scholastics. He belonged to the Dominican order and studied and taught at the Sorbonne, where he helped to edit one of the editions of Aquinas' *Summa Theologica* and of the *Summa* of St Antonino of Florence. From 1522 to 1546, he taught at the University of Salamanca.

Domingo de Soto (1495–1560), also a Dominican, studied at Alcala and under Vitoria in Paris. His treatise *De Iustitia et Iure* (Soto [1553] 1968) went through no fewer than 27 editions in 50 years and continues to exert significant influence (Grice-Hutchinson 1978, p. 95). Another of the early Hispanic Scholastics was Martin de Azpilcueta, 'Dr. Navarrus' (1493–1586). Regarded as one of the most eminent canon lawyers of his day, his *Manual de Confesores y Penitentes* (Azpilcueta 1556) was one of the most widely used spiritual

handbooks in the century following its publication. Azpilcueta was also of the Dominican order. Other important Dominican Scholastics include Domingo de Bañez (1528–1604), Tomás de Mercado (c. 1530–76), Francisco García (García 1525–75) and Pedro de Ledesma (Ledesma d. 1616). Franciscans Juan de Medina (1490–1546), Luis de Alcalá and Henrique de Villalobos (d. 1637) employed Scholastic sources and methods. The Augustinian bishop Miguel Salón (1538–1620) as well as Pedro de Aragón (Aragón 1545/6–92), Cristobal de Villalón, Luis Saravia de la Calle and Felipe de La Cruz added to the body of Late Scholastic thought. With the foundation of the Society of Jesus in 1540, Jesuit thinkers such as Luis de Molina (1535–1600), Juan de Mariana (1535–1624), Francisco Suárez (1548–1617), Juan de Salas (1553–1612), Leonardo Lessio (1554–1623), Juan de Lugo (1583–1660), Pedro de Oñate (1567–1646), Juan de Matienzo (1520–79) and Antonio de Escobar y Mendoza (1589–1669) made significant contributions.

The Latin American connection

Of the above, Tomás de Mercado (Mexico), Juan de Matienzo (Peru), and Pedro de Oñate (Lima and Cordoba, Argentina), spent most of their academic lives in Latin America. Luis López (1530–95, Chiapas and Guatemala); Bartolomé de Albornóz (Albornóz 1520–75/76) (Mexico), and Domingo Muriel (1718–95, Córdoba, Argentina) also taught at some of the 21 universities founded in Latin America during the first two centuries of colonization.

From the religious orders to the Scholastic scholars

Building on the Thomistic tradition, the Late Scholastics ascribed great importance to the justification of private property, stating that it derives from both eternal and natural law. Some of the early Scholastic authors had argued that things should be owned in common and had condemned those who possessed riches. The Late Scholastics rejected this condemnation, employing scriptural arguments and analysis of human action to prove their point. Those who opposed private property often quoted the passage that we quoted above describing the rich young man (Luke 18:18–25). Although many authors think that Jesus was condemning the possession of riches, the Late Scholastics indicated that this was not the correct interpretation. Citing Luke 14:26, where Jesus says, ‘If any man come to me without hating his father, mother, wife, children, brothers, sisters, yes and his own life too, he cannot be my disciple’, the Scholastics pointed out that this passage does not enjoin Christians to hate their fathers. Such doctrine would contradict the Fourth Commandment. Thomist and Scholastic interpretation of this passage is that entry to the kingdom of Heaven is denied to anyone who values things more than God. Aquinas wrote, ‘The rich man in question was criticized for thinking that external goods were his absolutely as if he had

not first received them from another, namely God' (*Summa Theologica*, II-II, qu. 66, art. 1, resp. 2.)

Those who opposed private property also cited Acts 2:44–7: 'The faithful all lived together and owned everything in common; they sold their goods and possessions and shared out the proceeds among themselves according to what each one needed'. As did Aquinas, the Late Scholastics recalled Augustine's condemnation of the teachings of the so-called 'Apostolics'. They declared that it was heresy to say that those who have property cannot enter the kingdom of heaven. Lessio noted that there are many passages in the Scriptures that state that possession is not a sin (Lessio 1605, Lib. 2, Cap. 5, Dub. 2, p. 41). Aquinas quotes Augustine in his *Summa Theologica*, II-II, qu. 66, art. 2, resp.:

> Augustine says: The people styled apostolic are those who arrogantly claimed this title for themselves because they refused to admit married folk or property owners to their fellowship, arguing from the model of the many monks and clerics in the Catholic Church (*De Haeresibus* 40). But such people are heretics because they cut themselves off from the Church by alleging that those who, unlike themselves, marry and own property have no hope for salvation.

Pedro de Aragón explained that if we suppose that it is better for certain men (for example, members of religious orders) to possess goods in common, it does not follow that the same can be said for all human beings in general (Aragón 1596, pp. 110–11). By the same token, one might conclude that since priests must remain celibate, no one should marry.

In addition to scriptural references, the Medieval Schoolmen offered utilitarian arguments to show that goods that are privately owned are better used. Domingo de Soto criticized common ownership from an Aristotelian and Thomist perspective. Stating that it is impossible to achieve abundance in a common property system, he suggested three possible arrangements: (a) land that is privately owned while its produce is commonly shared; (b) commonly owned land whose produce is privately enjoyed; and (c) common possession of both the land and its fruits. Soto admitted that each of these systems has its drawbacks.

In the first case, disputes will arise. The rewards of labour will be unequal. Those who own more land will have to work more, while the fruits of their labour will be distributed to all equally according to need. They will resent receiving less for working more (Soto [1553] 1968, bk. IV, qu. 3, fol. 105–6). With common ownership of land involving private ownership of fruits, 'everyone will expect the others to do the work'. Since people's love for their own goods is strong, 'the distribution of goods will cause great envy' (ibid.). Similar problems would arise if both the land and its fruits were commonly owned: 'Each worker will try to appropriate as many goods as possible, and

given the way human beings desire riches, everyone will behave in the same fashion. The peace, tranquillity and friendship sought by the philosophers will thus inevitably be subverted' (ibid.)[16]

When goods are commonly owned, the orderly society and a peaceful division of labour are impossible. Since no one will be willing to accept the more dangerous jobs, society will forfeit tranquillity. In addition to their economic arguments in favour of private property, the Late Scholastics also cited moral reasons. Soto wrote that in a context of commonly owned goods, the virtue of liberality would disappear, since 'those who own nothing cannot be liberal'.

Tomás de Mercado also acknowledged the existence of self-interest and the greater care that humans generally exercise in relation to their own property. Realizing that economic goods are scarce, Mercado espoused private property as an efficient method of reducing – if not overcoming – scarcity, 'We cannot find a person who does not favour his own interests or who does not prefer to furnish his home rather than that of the republic. We can see that privately owned property flourishes, while city- and council-owned property suffers from inadequate care and worse management' (Mercado 1571, bk. II, chap. II, fol. 18–19).

Juan de Mariana also referred to the relationship between self-interest and the careful use of economic goods. With a high degree of self-criticism, he cited the example of the poor use the Jesuits made of the things they owned in common:

> We are too extravagant. Our cassocks are made of black woollen cloth and we are supplied everything in common, from the littlest to the biggest items: papers, ink, books and our provisions for journeys [*viaticum*]. Certainly it is natural for people to spend much more when they are supplied in common than when they have to obtain things on their own. The extent of our common expenses is unbelievable. (Mariana [c.1600] 1950, p. 604)[17]

Luis de Molina included many passages favouring private property in his *De Iustitia et Iure*. If things were held in common, he said, the powerful would inevitably exploit the weak. No one would be interested in serving the public good, and no one would agree to work in those jobs that require greater effort (Molina 1614, col. 100–101, Treatise 2, disputation 20). According to Molina, private property may have existed even before original sin, since in that state, men could agree by common consent to divide the goods of the earth. The commandment 'thou shall not steal' implies that the division of goods does not pervert natural law (ibid., col. 102, Treatise 2, disputation 20).[18] All the Late Scholastic authors granted considerable importance to the moral use of goods that private property allows. 'Alms should be given from private goods and not from the common ones' (Vitoria 1934,

Summa Theologica II-II, qu. 66, art. 2, p. 324). The virtues of charity liberality, hospitality and generosity would all become impossible in a world without private property.

Analysing the Late Scholastics, Joseph Schumpeter wrote:

> The individualist and utilitarian streak and the emphasis upon a rationally perceived Public Good run through the whole sociology of St. Thomas. One example will suffice: the most important one, the theory of property. (Schumpeter 1954, p. 92)

> [Late Scholastic] economic sociology, especially their theory of property, continued to treat temporal institutions as utilitarian devices that were to be explained – or, 'justified,' – by considerations of social expediency centering in the concept of the Public Good. (ibid., p. 96)

It is not surprising that these authors employed utilitarian arguments, especially since they preceded them with demonstrations that the division of goods is in accordance with natural law. For the Late Scholastics, however, the division of external things was a matter of *ius gentium*, which stems from different principles from the ones on which natural law is based (Soto [1553], bk. IV, qu. III, fol. 105). 'We know that the fields are not going to be efficiently tilled in common ownership and that there will not be peace in the republic, so we see that it is convenient to undertake the division of goods' (ibid.)

After repeating similar arguments, Antonio de Escobar y Mendoza explained that nearly all people, except the most savage, have given their consent to the division of property because goods are better administered in private hands (Escobar y Mendoza 1662, ch. III, p. 4).

Subsurface property

Aquinas (*Summa Theologica* II-II, q. 66, art. 5, resp. 2) and many of his disciples discussed the subject of ownership in reference to things found both on the surface of the earth and underneath it. Their analyses and conclusions are important for contemporary economic policies because current legislation in many countries provides for different treatments of 'surface property' and 'underground property'. Pedro de Ledesma, following St Antonino's reasoning, remarked that those things that have never had an owner 'belong to the one who finds them, and the one who finds them does not commit theft by keeping them' (Ledesma 1614, p. 443).[19] According to Ledesma, the finder has a natural right to appropriate such goods. He also recognized that in many kingdoms there were laws that overrode this right.

Those things that at one time had a proprietor, such as treasures, may, in certain circumstances, belong to the one who found them. Sometimes the

finder may not keep the treasure, as, for instance, when the owner's family knew where the treasure was hidden. Miguel Salón remarked that the nature of the circumstances can affect the question of ownership. Salón criticized the Spaniards who appropriated Aztec and Inca treasures, describing this as simple robbery and declaring that these treasures should be restored to their real owners. By natural right, however, any treasures found in no-man's land belong to the finder (Salón [1581] 1591, col. 1298).[20] Salón specified that the same rule applied when someone found a treasure on his own property.

The majority of the Thomist authors used Matthew 13:44 to prove the right to a treasure: 'The kingdom of heaven is like treasure hidden in a field which someone has found; he hides it again, goes off happy, sells everything he owns and buys the field'. Both Salón and Ledesma employed this argument.

If natural law says that a treasure belongs to its finder, either totally (as when the treasure was found on one's property or in no-man's land) or partially, it is logical to conclude that everything placed by nature under the earth's surface reasonably belongs to the owner of the surface. The Scholastics cited the examples of metal and mineral deposits, especially silver and gold. Salón stated explicitly that 'the minerals and gold and silver deposits, as well as any other metal in its natural state, belong to the owner of the land and are for his benefit' (Salón [1581] 1591, col. 1307).[21] More than a century later, the Late Scholastic author P. Gabriel Antoine (1678–1743) judged that:

> [S]tones, coal, clay, sand, iron mines, lead, which are found under someone's land, belong to the owner of the land. In effect they are part of the land, for it does not consist solely of the surface alone, but of its entire depth all the way to the earth's centre, and here is where we can find these fruits. And the same can be said of metal deposits. (Antoine 1774, First part, ch. II, qu. VI, p. 369)

For Soto, anyone who found gold or other metals under an ownerless plot of land had ownership rights to those minerals. Taxes would only be 'natural' if the metal was found in a property belonging to the prince. The only exception Soto could find was in a case of grave public need, but even then he declared this an insufficient reason. Yet for metals, Soto accepted a 20 per cent tax, or what was called the 'metallic fifth' (Soto [1553] (1969), Book V, a. III, f. 151).

Ownership and use of property

Quoting Conradus Summenhart, Francisco de Vitoria defined domain (*dominium*) as the faculty to use an object according to reasonably established laws. People can use things although they are not the owners. In this sense, domain and use are distinct. Whenever people have perfect domain over a good, they can use this good as they please, even to the extent of destroying it. As Villalobos pointed out, 'Domain has to do with the substance of the

thing; and so, the one who owns it can sell it, transfer it and if he wants, destroy it' (Villalobos 1632, p. 126).

Among the uses of property, transfers of domain are essential for economics. Exchanges are, in essence, a transfer of domain. Soto acknowledged that 'there is nothing so much in agreement with natural justice as to enact the will of a man who wishes to transfer the domain [property] of his goods' (Soto [1553] 1968, bk IV, qu. V, fol. 110). 'Any person has the natural right to donate or transfer the things he legally owns in any way he wants'. Soto added that if man can be a property owner because he has free will, by this same free will he can transfer his domain to anyone else (ibid.).[22]

As all things have been created for man, he may use them as he pleases. Moreover, the ownership of something consists of the faculty and the right to use that thing in every way permitted by law, such as donating, transferring, selling or consuming it in any manner. Despite this natural right, Soto declared that the law may restrain the will of the owner and even deprive him of his good against his will. Although man is a social being and he will therefore find it advantageous to live in society, the republic needs an authority, and the main function of public authority is to defend the republic and to administer justice. To fulfil its duty, authority has to supervise the use that young people make of their goods until they reach the age of full reason. Second, Soto declared that some goods must be used to support authority (in the form of taxes). Third, authority has the duty to punish crimes. One form of punishment is depriving the guilty party of his goods. Other restrictions on the use of property refer to the use of ecclesiastical goods.

Extreme need

As in other topics, the Scholastic analysis on extreme need is influenced by the writings of Aquinas, who points out that:

> [When] the need be so manifest and urgent, that it is evident that the present need must be remedied by whatever means be at hand (for instance when a person is in some imminent danger, and there is no other possible remedy), then it is lawful for a man to succor his own need by means of another's property, by taking it either openly or secretly: nor is this properly speaking theft or robbery. (*Summa Theologica* II-II qu. 66, art. 7)

Aquinas stressed that we confront an economic problem: on one side there is a large demand for help from the indigent, on the other, we have scarce means to satisfy this need. One has to choose who will receive help, and that should be left to the free will of the owners. Aquinas continues: 'Reply to Objection 2. It is not theft, properly speaking, to take secretly and use another's property in a case of extreme need: because that which he takes for the support of his life becomes his own property by reason of that need'. One could also take other

people's property to help the neighbour in extreme need. In other words, taking goods which one does not own can only be justified when there is no other way to avoid the death of the person. If a small plane falls in a plantation in the middle of a jungle, and some of the wounded passengers cannot procure the necessary fruits, it is justifiable, in Thomistic thinking, that a person will take those fruits to help the person in need.

Martín de Azpilcueta, also analysed the topic of extreme need, and his conclusions are even more respectful of private property:

> No one is obliged to donate anything to he who is in extreme need: because it suffices that he lends him what is necessary to liberate him from it, and the person in need has no right to take more of the neighbour's estate than its owner, and it is enough, if there is a need, that he takes it as a loan and not as his own. (Azpilcueta 1556, p. 206)

It is not proven either 'that extreme need makes the needy absolute owner of the neighbour's goods, it only gives them a right to use them if it is necessary to escape the need … extreme need by itself does not make one a lord over the neighbour's goods without a duty to restitute'.

Those of us who assign great importance to private property, can agree with Azpilcueta that he 'who takes something in extreme need, is obliged to make restitution when he has a chance: independently if he has goods in another place or not, and even if he had or had not consumed the goods' (ibid., p. 207). Due to space constraints, we have presented here mostly arguments of a theological and juridical nature. Nevertheless, the influence of these authors in the culture was perhaps even greater through their sermons and confessions. They told wonderful stories in defence of property and repeated popular sayings such as 'a donkey owned by many wolves is soon eaten' (Albornóz 1573, p. 75).

One could conclude with Bede Jarrett's analysis that:

> [For these Schoolmen] the right to property was an absolute right which no circumstances could ever invalidate. Even in case of necessity, when individual property might be lawfully seized or distrained – in the name of another's hunger or of the common good – yet the owner's right to property remained and endured. The right was inviolable even when the exercise of the right might have to be curtailed. (Jarrett 1942, p. 123)

Summary

Late Scholastic thought provides several arguments in favour of private property:

1. Private property helps to ensure justice. Evil exists because men are sinners. If goods were commonly owned, it would be the evil men 'and

even the thieves and misers' (Vitoria [c. 1530] 1934, p. 235) who would profit most, since they would take more from the barn and put less into it. Good men, on the other hand, would contribute more and profit less. The fact that the most immoral people dominate society represents a harmful element and a distortion of natural order.

2. Private property is useful for the preservation of peace and harmony among men. Whenever goods are held in common, disputes are inevitable.
3. Privately owned productive goods are more fruitful because it is natural for men to take better care of what is theirs than of what belongs to everybody; hence the medieval proverb, 'A donkey owned by many wolves is soon eaten'.
4. Private property is convenient for maintaining order in society, and it promotes free social cooperation. If everything were held in common, people would refuse to perform the less pleasant jobs.
5. No man (not even a priest) can detach himself from temporal goods. Original sin brings with it the problem of scarcity, which is the source of economic problems (that is, the difference between unlimited needs and limited resources). 'This participation and division is so necessary because of our own weakness and misery. These principles must apply even to members of religious orders who choose poverty in a desire to imitate original innocence. The prelates of such orders must distribute vestments, books, papers and other items so that the priests will make good use of some and those in need can use the rest' (Mercado 1569, fol. 18).

Creating a culture of private property

> In questions of right we must diligently guard against attributing too much power to the State. (Matteo Liberatore 1891, p. 181)

The contributions from the Late Scholastics did not pass unnoticed to important lay authors from the United Kingdom. James Mackintosh (1765–1832) wrote:

> Both he [Soto] and his master Victoria [sic] deserve to be had in everlasting remembrance, for the part which they took on behalf of the natives of America and of Africa, against the rapacity and cruelty of the Spaniards. Victoria pronounced war against the Americans, for their vices or for their paganism, to be unjust.
>
> Soto was the authority chiefly consulted by Charles V, on occasion of the conference held before him at Valladolid, in 1542, between Sepúlveda, an advocate of the Spanish colonists, and Las Casas, the champion of the unhappy Americans, of

> which the result was a very imperfect edict of reformation of 1543. This, though it contained little more than a recognition of the principle of justice, almost excited a rebellion in Mexico. (Mackintosh 1851, p. 23)[23]

Lord Acton also praised them: 'The greater part of the political ideas of Milton, Locke, and Rousseau, may be found in the ponderous Latin of Jesuits who were subjects of the Spanish Crown, of Lessius, Molina, Mariana, and Suarez' (Acton [1907] 1993, p. 82).[24]

From the mid-seventeenth century to the early nineteenth century there were few radical departures or contributions to the Christian approach to private property. Major contributions came from lay authors, such as John Locke and Francis Hutcheson, which had such a great influence on the works of Adam Smith and David Hume. Gaetano Filangieri (1752–88), Abbé Galiani (1728–87) and Abbé Genovesi (1712–69) in Italy, were influential not only in Italy but also in Latin America, where they were some of the most widely read authors. Anne Robert Jacques Turgot (1727–81) and Abbé Etienne Bonnot de Condillac (1715–80) in France, were also great champions of private property within a Christian perspective. But the author who had the greatest impact in preparing the ground for a rule of law respectful of private property was John Locke (1632–1704).

Locke was identified by Joseph Schumpeter (*History of Economic Analysis*) as among the 'Protestant Scholastics' of whom Locke's forerunners were Richard Hooker (1554–1600), Hugo Grotius (1583–1645) and Samuel Pufendorf (1632–94). Locke, a student and medical teacher at Oxford, had been influenced by the Scholastic Aristotelianism of Oxford. Cambridge Platonism contributed to Locke's critique of Thomas Hobbes's political views against which Locke wrote his *Second Treatise of Civil Government* (1690).

Another major English opponent of Hobbes's attack on individual rights was Chief Justice Matthew Hale (1609–76) in his *Treatise of the Nature of Lawes in Generall* (1660) showing that in the state of nature men had an exclusive right to private property. The Cambridge Platonist Richard Cumberland (1631–1718) in *De Legisbus Naturae Disquisitio Philosophica* (1672), basing his critique of Hobbes on Grotius, showed that in the state of nature property rights existed, and preceded any necessity for a formal contract of all men. Cumberland's work appeared the same year as Samuel Pufendorf's *De Jure Naturae et Gentium* (1672) and Pufendorf commended Cumberland's book in the next edition (Tuck 1979, p. 167).

Locke's *Two Treatises of Government* were written in 1680 and published after the Glorious Revolution of 1688. The *First Treatise* aimed at demolishing the arguments of Sir Robert Filmer's *Patriarcha* (London, 1680), written in the 1640s to defend the Absolutism of King Charles I. Filmer argued from the Book of Genesis that God had granted to Adam as the first man the absolute

government over all men and their lands. He did not explain how one would know that the Scottish Stuart family or the Welsh Tudor family were the inheritors of Adam's absolute ownership of men and wealth in England. Filmer believed that Absolutist rule over men and land was necessary given the fallacy of the concept that all land was owned by all mankind in common. Common ownership would mean the destruction of mankind for Filmer since everyone would die before anyone could gain unanimous consent of mankind to one's taking a drink of water or an apple from the common pool.

Locke's response to Filmer is Christian rather than Aristotelian:

> God did not give Adam absolute and arbitrary power over the world or over his children.
>
> God is a rational God who gives any of his agents only such power as they need to fulfill the purposes for which he gives them that power. The violence of men's wills is such that to grant them absolute power is sure to corrupt them; what Adam gained from God was the right to use nature and direct his offspring so that they might flourish as God intended. (Ryan 1984, p. 15)

In the *Second Treatise*, Locke moves to explain his alternative to Filmer and Hobbes. Locke posits that private property exists, as does marriage and the family, antecedent to the invention of government. Private property and the family belong to the state of nature, and government is subsequently superimposed. Locke shows that the purpose for the imposition of government is for the protection of private property. Locke needed to present a rational explanation of private property in order to defeat Filmer's argument for Absolutist government.

Locke accepted the idea that private property was just if others could similarly create private property by mixing their labour with the natural resource. Locke had made a serious study of the immense travel and voyage literature written by those who had visited the non-European peoples around the world. He was a comparative anthropologist:

> If the close ties between politics, economics, sociology, and anthropology are kept in mind, it may be appropriate here to second an observation previously made by Harrison and Laslett; namely, that Locke possessed one of the finest collections of voyage and travel books in the seventeenth century. He was forever noting in his commonplace books the political, economic, and social customs recorded by travelers to non-European countries. Moreover, as a political adviser to others, a polemicist on economic problems, and a shaper of policy in his own right through his membership of the Board of Trade, Locke was one of the few persons capable of effecting changes in the economic thought and structure of his society. (Ashcroft 1990, p. 242)[25]

As a leading economist, at a time when Sir Isaac Newton (1642–1727) was the Master of the Mint, Locke wrote reports regarding money. He concluded

that the provision that private property was accompanied by a need that others should have the opportunity to mix their labour with the land was achieved by the evolution of money. Money permits the storage of value in material resources which do not expire and do not spoil. Money permits the exchange of material resources and services, and their storage and saving. A money economy creates a more perfect system for all to share in God's creation. Alan Ryan notes:

> The reason why inequality and the occupation of all the vacant land do not violate anyone else's rights is that what the appropriator has to do is leave enough and as good for others, not in the sense of leaving as much land for others, but in the sense of leaving others just as able as they were before to get what Locke terms a 'living'. The day-laborer who has no land none the less gets a good bargain from the process whereby money and inequality have advanced together, for he lives, lodges and is clad better than the king of an Indian tribe in the empty wastes of inland America. (Ryan 1984, pp. 17–18)

Locke presented a worldview which is more Christian than Aristotelian. But, most medieval authors had been seeking to adhere to the Aristotelian worldview, creating confusion in their analysis. The Christian Fathers had left a legacy to the West of the 'Idea of Progress'. Greek thinkers had begun to express the concept of the natural growth of knowledge in time and thus the advance or progress of the human condition. The Christian Fathers expanded and deepened this concept and gave it a spiritual force. Robert Nisbet said:

> I refer to such attributes as the vision of the unity of all mankind, the role of historical necessity, the image of progress as the unfolding through the long ages of a design present from the very beginning of man's history, and far from the least, a confidence in the future that would become steadily greater and also more *this*-worldly in orientation as compared with *next*-worldly. (Nisbet 1980, p. 47)

Nisbet's focus on Christian worldliness of the Church Fathers is based in the theology of the Incarnation (God becoming Man) and of the *Imago Dei* (man created in the image and likeness of God). An important source of Nisbet's thinking was based upon the work of the Fordham University historian, Gerhart B. Ladner, whom he quotes as authoritative:

> All the great social concepts of Christianity, be it the Kingdom of God or the Communion of Saints, the Church, or the City of God, were conceived as immanent in this world, as struggling, and time-bound, and, at the same time, as transcendent, as invisible, as triumphant, as eternal in the world to come. There was, then, a Church as the community of the faithful on earth and a Church as the congregation of the heavenly city consisting of its citizens, that is to say, the saints and the angels. It is one of the great paradoxes of Christianity that such a concept

> as the Church can appear under these two aspects and nevertheless be ultimately one. (Nisbet 1980, p. 49; Ladner 1959)

The study of the implications of the *Imago Dei* on the understanding of the human person led to the theology of the individual and freedom of the will in Western Christian thought. Nisbet states:

> Gerhart B. Ladner in his magisterial *The Idea of Reform: Its Impact on Christian Thought and Action in the Age of the Fathers* has given us what is by all odds the most detailed and illuminating account of early Christian preoccupation with reformation and advancement of their own society. Ladner writes: 'In modern times the term and idea of reform are applied to the renewal and intended improvement of many things, more often however of social entities and institutions than individuals. The origins of the Christian idea of reform on the contrary are related to the core of evangelical and Pauline doctrine on the human person: to the experience of its newness in Christ. … And yet, in spite of the personal, individual character of the Christian idea of reform it became effective as a *supra-individual force* at a relatively early date.' (italics added) (Nisbet 1980, p. 57; 1979, pp. 7–37)

Although the medieval period had experienced a vast increase in technology and productivity compared to the stasis of the ancient world, the scholars' rediscovery of Aristotle led them to narrow their focus to the search for the static in the Greek world.

The Aristotelian ideal was of the citizen having a reasonable income with much leisure to engage in politics and militia training. Increased productivity was a danger to the leisure of politics and militia training. The merchant or the craftsman could not be a citizen because he did not have the leisure for politics and militia training. Aristotle's ideal is the avoidance of work, and those who work like merchants or craftsmen are demeaned creatures. Locke shares the Christian insight of the holiness of work, captured in the Benedictine ideals: to work is to pray. Private property encourages the hard work which is an expression of the gratitude to God for His gifts to us. God created the world unowned, open to the conversion into private property for man's survival. God would not have given the world to man without a means to use the world, leaving men to starve or to commit theft of taking goods from the common pool (Ryan 1984, pp. 30–31).

Locke sees the necessity of private property as

> [The means to reward men for 'taking pains'] for taking the intrinsically disagreeable effort required to make things serve human purposes. Work is certainly important; and working is morally desirable, both as a matter of individual survival and as a matter of social improvement – unless we work we shall starve, and unless we work well, we shall create none of the surplus on which social progress depends. (ibid., p. 28)

Yet, the medieval scholars were oblivious to the riot of application of technology surrounding them. They followed Aristotle in opposing the taking of interest, while the Italian merchants found many forms in which investment of capital and its return of interest were acceptable. Locke was seeking to bypass Aristotelian economics in favour of a Christian foundation based on St Paul's and the Church Fathers' spirit of individual flourishing and progress.[26]

Ryan sees Locke's individualism as essentially Christian, but freed from the static economy of the Christianized Aristotle and accepting the dynamic spirit of the Early Church Fathers in a Reformation fashion. Perhaps, Locke's mindset of the Early Christians led him to reject the state's intrusion on the individual's rights, as in life, liberty and property, in the context of the state's 'incompetence to achieve our spiritual good, and in terms of God's unwillingness to receive a forced submission' (ibid., pp. 23–5).

Locke devoted Chapter V in the *Second Treatise* to reason's demonstrating that men are to be defended in their possession of property. Locke begins with the unassailable premise that each man has property in his own person. Self-ownership was common among moralists of the seventeenth century and was introduced into northern Europe by Hugo Grotius. Locke explained:

> [E]veryman has a property in his Person. This no body has any right to but himself. The labour of his body, and the work of his hands, we may say, are properly his. Whatsoever then he removes out of the state that nature hath provided, and left it in, he hath mixed his labour with, and joined to it something that is his own, and thereby makes it his property. It being by him removed from the common state nature placed it in, it hath by this labour something annexed to it, that excludes the common right of other men. (Locke [1690] 1970, pp. 305–6)

Locke not only presents the moral justification for private property but also the utilitarian benefits to men. Common land which no one is motivated to utilise provides almost no benefits to men. Nature which is not improved upon provides very little of use to men.

> I think it will be but a very modest computation to say, that of the products of the earth useful to the life of man, 9/10 are the effects of labour: nay, if we will rightly estimate things as they come to out use, an cast up the several expenses about them, what in them is purely owing to nature, and what to labour, we shall find, that in most of them 99/100 are wholly to be put on the account of labour. (Locke 1970, pp. 314–15)

Men created in the image and likeness of God must co-create with God by creating private property by men's intelligence and planning. It would be a sin against God for a man to allow the earth's goods to spoil. The limit on possession was not defined by quantity, but by the consequences of owner-

ship – that nothing spoils. Locke believes that in the natural activity of trading, the most durable good becomes accepted as money.

But, the invention of money meant that goods could be stored in value without spoilage:

> Again, if he would give his nuts for a piece of metal, pleased with its color: or exchange his sheep for shells, or wool for a sparkling pebble or a diamond, and keep those by him all his life, he invaded not the right of others, he might heap up as much of those durable things as he pleased; the exceeding of the bounds of his just property not lying in the largeness of his possessions, but the perishing of any thing uselessly in it. And thus came in the use of money, some lasting thing that men might keep without spoiling, and that by mutual consent men would take in exchange for the truly useful, but perishable supports of life. (ibid., pp. 318–19)

Christian thought and the attacks on private property

Catholic thought began to concentrate again on private property as a response to the surge in collectivist thought. Father Jaime Balmés (1810–48) was one of the first to analyse the dangers of socialism and its attack on private property. A Catalonian who spent a large portion of his short life in France, devoted a series of articles to the issue. Apart from repeating the economic and ethical arguments in defence of a social order based on private property he also uses historical examples. Criticizing Robert Owen he wrote 'What happened in [the socialist experiment of] New Harmony is not an exceptional case but an example of what would naturally occur in all periods and countries'. There could be minor differences according to differences in the people but the end result would be the same. When devising reforms one has to take into account 'men as they are, not as we would like them to be' (Balmés [1844] 1925, pp. 265–6). In his philosophical texts, Balmés also had major chapters in defence of private property and cautioning about the calamities that the world was going to face due to the growth in socialist sects.[27]

Matteo Liberatore (1810–92) was born in Salerno, Italy. He was a Jesuit and one of his many efforts was the founding of the journal *Civiltà Cattolica* in 1850. Liberatore embarked on a programme to revive the study of St Thomas Aquinas and wrote many books and papers on ecclesiastical and social matters. His volume on economics was targeted to the educated layperson. His views served as the basis for the economic aspects of *Rerum Novarum* (1891): the first great social encyclical. This encyclical has been the point of reference for *Quadragesimo Anno* (1931), written to commemorate its fortieth anniversary, and *Centesimus Annus* (1991), which was written to celebrate its centennial.

It is on the topic of property where Liberatore left the most lasting legacy: 'Property means exclusive possession of a thing with power to dispose of it at will' (Liberatore 1891, p. 269). Antonio Rosmini (1797–1855) also had great

insights about private property and the human person. He described it as a sphere around the person, with the person at the core, '*una sfera intorno alla persona di cui la persona è il centro*'(Rosmini [1845] 1967, p. 154).

Socialist criticisms are against 'nature commands' as private property is necessary to secure peace, an abundant production and help purvey for one's own future and that of our children (Liberatore 1891, p. 330). Consistent with all Christian writers, Liberatore agrees that property owners are obliged by conscience to give their superfluous wealth to the deserving poor. In certain occasions the state has the right to regulate but not abolish private property.

Liberatore added that the views of the Late Scholastics, who argued that private property belonged to positive and human law, could create misunderstandings. During Liberatore's times, human positive law was understood as that which depended solely on the will of the legislators. To continue with the old terminology would be imprudent and dangerous and favour the socialists who would 'play tricks of sophistry and confuse the question, by maintaining that because in modern language the right of having property is not given by nature, but by man, therefore private property may be abolished' (ibid., p. 137). Liberatore's preference for regarding private property as a natural right is in part because he was convinced that it is 'better to use phrases not liable to false interpretation but easily understood' (ibid., p. 137). He concludes: 'Now-a-days the question of property has left the School to the streets, and writers, therefore, should take the greatest care to avoid words that may be misused' (ibid., p. 138). Liberatore does not disagree with the Scholastic notions, he rather adapts them to the language of his times.

On certain occasions the state has the right to regulate but not abolish private property. According to Liberatore, not even the consent of all the nations could justify the abolition of private property. Common ownership could be imposed only by the unanimous consent of individuals (for example, shipwrecked people on an island) but, according to him, the children and grandchildren would not be obliged to obey because they 'receive the right of having property, from nature' and not from their progenitors (ibid., p. 134).

> [The socialist system is] evidently absurd; for it means that all the individual rights and powers of the subjects ought to be absorbed by the State. No man of common sense can seriously entertain such a notion, and therefore to speak of it further would only be a waste of time. (ibid., p. 184):

> State socialism, by which the government makes itself the arbitrary master of the production and distribution of the national wealth, certainly is abominable. If society suffers so much now, as in fact it does, from bureaucrats absorption in the administrative order, what will it be when the economic order has fallen into its clutches? (ibid., p. 204)

Pope Leo XIII (1810–1903) adopted many of Liberatore's views on property rights. This prompted Jacob Viner to write:

> [The] sceptical attitudes at the end of the Middle Ages towards the doctrine of a natural-law foundation of the institution of private property indicate that St. Thomas' use of even a quasi natural-law justification of private property was by no means universally accepted. It seemed to some, therefore, that Pope Leo XIII broke sharply with traditional doctrine when he went beyond St. Thomas in his campaign against nineteenth-century socialist doctrine, by proclaiming as an integral part of natural law the right of private property. (cited in Melitz and Winch 1978, pp. 70–71)

Luigi Taparelli D'Azeglio (1793–1862) went further by analysing the importance not only of ownership but of secure ownership. Abundance of wealth is not enough for the happiness of a people, secure ownership is more important.[28]

Tranquillity, stability of contracts, grow in proportion to the development of commerce in society, it is the nature of society and 'is born and grows under the tutelage, under the mantle of law and justice' (Taparelli, 1855, #979). When contracts are not respected, society lives in permanent anxiety and each owner and trader will always have a doubt of himself and of being dispossessed due to a defect in the old contracts. That is why, as others before, Taparelli favours *prescription* as a way to solve the conflict between a private right and the 'right of society to peace and progressive wealth' (ibid.), which would suffer greatly by uncertainty. The private right yields to the social right for reasons of evidence, generality and importance. The minor right yields to the major.

Liberatore was a great defender of inheritance rights and devoted an entire chapter of his books on economics to it. For him it is inconsistent to believe in property rights and at the same time not defend the right of inheritance. He defended inheritance on economic, legal and political grounds, and stressed the importance of inheritance to protect the family. Here he quotes Taparelli at length, in his defence of inheritance as an essential means to protect domestic–social unity (Liberatore 1891, p. 144).

Oswald von Nell-Breuning was born at Trier in 1890. A Jesuit, he was ordained in 1921 and from 1927 onwards became professor of moral theology and canon law on the university faculty of theology of Sankt-Goergen, Frankfurt-am-Main. Nell-Breuning was the main writer of the draft of the encyclical *Quadragesimo Anno*. He wrote a book to further explain the economic views of Pope Pius XI. All his quotations appearing in this chapter come from that work, *Reorganization of Social Economy* (1937).

Nell-Breuning also defends inheritance rights and shows the continuity in Catholic thought, exemplified by Pius XI's use of key passages of *Rerum Novarum* 'Man is older than the state' (*Rerum Novarum* 6), and 'the domestic

household is antecedent as well in idea as in fact to the gathering of men into a community' (*Quadragesimo Anno* 42 citing *Rerum Novarum* 10).

'God created all worldly goods for humanity as a whole, and they must, under all conditions, be kept for the purpose established by God's ordinance. The institution of personal property is merely the *means* to secure in an orderly way, the fulfilment of the *end* of the ordinance' (Nell-Breuning 1937, pp. 96–7). Nell-Breuning quoted Benedict XV who praises the Roman law as the 'glorious memorial of ancient wisdom which is justly called reason in writing' (*Romanorum ius, insigne veteris sapientiae monumentum quod ratio scripta est merito nuncupatum*) (ibid., p. 99). 'In using the Roman legal-property concept, moral theology has chosen an efficient scientific instrument. And in science not the tool but its proper use is important' (ibid., p. 101).

Nell-Breuning concludes his analysis of property rights with a section entitled 'Actual abuse of one's own property rare'. How do we know if we are using property within our rights? 'The observance of this one virtue [commutative justice] will assure the owner of a clear conscience with regard to the limits of his property right. All other natural and Christian virtues come into play even before the virtue of commutative justice' (ibid., p. 102). He then ends the section by saying: 'In practical life, people as a rule demonstrate a rather high capacity of discernment, and only in a few difficult borderline cases at the most, will they be doubtful whether the owner has made a bad and morally disorderly use, or rather abuse, of his property right, or whether he lacked legal rights altogether, and is thus guilty of violation of law'.

Income is defined as 'the fruits of our ability to earn a livelihood as well as other returns … permanent property, especially as a source of income, must be preserved as far as possible, must perhaps even be increased, while income is used primarily to supply the means for consumption' (ibid., p. 114).

Nell-Breuning argues that socialists frequently find refuge 'in the principle of envy expressed in the question: Why should others be better off than myself? This is psychologically effective, but fundamentally wholly worthless. This envy, the desire to have what others have, does indeed create a tremendous impetus in all those who are after the material things of life' (ibid., p. 290)

Most market-oriented economists have agreed with Karl Marx that state ownership of the means of production is the essence of the socialist system. This is not the case with Nell-Breuning for whom the essence of socialism is its anti-Christian philosophy and that it must be 'anti-God'. Socialism can even adopt private property and still be socialism (ibid., p. 305). The encyclical stated that 'No one can be at the same time a sincere Catholic and a true Socialist' (Quadragesimo Anno, 120). Nell-Breuning acknowledges that there 'may be Socialists who are also religious, or "Christian" according to their

own idea' but that they cannot be Christian Socialists because it is a contradiction in terms.

The Americas

Religious views on private property

Writing almost a century ago, Archbishop John Ireland (1838–1918) stated that 'the Papacy is recognized as the first and greatest moral power in the world. … To-day, as seldom before, the Papacy is prominent in the world as the religious and spiritual teacher of mankind' (Ireland 1905, p. 407). Ireland, as well as Cardinal James Gibbons (1834–1921), did their best to promote recognition by the papacy of the virtues of the American private property system.

Bishop Ireland correctly describes the thought of 'his' Pope Leo XIII:

> There must always be among men an unequal distribution of the possessions of earth; for the gifts of mind and body, through which these possessions are acquired, are unequal in men. The rights of property are sacred and cannot be violated; they who wrest to themselves the property of others are robbers, and, together with idolators and adulterers, are excluded from the Kingdom of Heaven. Nor can the State arrogate to itself the rights of men to private ownership, for man is older than the State, and private ownership is nature's own institution. (ibid., p. 412)

In his essay 'The Catholic Church and civil society', Archbishop Ireland quotes Charles Montalembert's (1810–70) views on the Middle Ages, a period 'bristling with liberty'. 'The spirit of resistance', Montalembert continues, 'the sentiment of individual right, penetrated it entirely; and it is this which always and everywhere constitutes the essence of freedom'. Ireland argued:

> Feudalism was at the time strongly entrenched in Europe, and opposed powerful obstacles to the development of liberty. The Church alone was capable of resisting its influences. 'If the Christian Church had not existed', says Guizot, 'the entire world would have been delivered up to mere material force. The Church alone exercised a moral power'. Hume himself writes that without the Papacy 'all Europe would have fallen very early into one or many caliphates, and would have submitted infallibly and disgracefully to Turkish sway and to Oriental oppression and stupefaction'. (ibid., p. 57)

Thanks in part to the strong presence of the Church and its views on limited power, 'the relative security of property rights that the medieval lords had secured was not extinguished, as in the absolutist states of the ancient civilizations' (Lal 1998, p. 73). Deepak Lal quotes Perry Anderson who wrote that the age that followed the late Middle Ages in which 'absolutist' public authority was imposed was also simultaneously the age in which 'absolute'

private property was progressively consolidated. 'It was this momentous social difference which separated the Bourbon, Habsburg, Tudor or Vasa monarchies from any Sultanate, Empire or Shogunate outside Europe' (ibid., p. 73). Very few societies were able to 'tie the hands of predatory governments' and only when government is limited can one speak of private property.

The right to depose the king was the essential right that limited authority. The greatest theologians of the Catholic Church agreed that that power resided in the people. This right had to be exercised very carefully '*Oportebit autem ut de tali jure, vel antiques et certis instrumentis, vel immemorabili consuetudine sufficienter constet*' (Only when based on antique and tried documents or a common law from time immemorial). Suárez mentions that Bellarmine took his views of limited power from Martín de Azpilcueta (the famous Dr Navarrus). Juan de Mariana expressed the same views and, like Suárez and Bellarmine, also taught at the College of Rome during the second half of the sixteenth century.

The American culture received praise from Ireland:

> It is a truth, which Americans do not fail to grasp, that, as Burke said, 'men have equal rights, but not to equal things'. Americans will not, in the hope of ulterior results, be willing to become parts of a vast machine, in which each one is but a link in a chain, or a cog in a wheel, without power of self-assertion. State socialism, even if it cloaks itself under the name of liberty, is in reality the veriest despotisms, and is radically opposed to the American mind and heart. (Ireland 1905, p. 209)

South of the border, in Latin America, Jean Jacques Rousseau was very popular, but not his ideas on property and religion. In Chile, Father Tadeo Silva was a great defender of property rights and a critic of Rousseau. So was Camilo Henriquez who even used Rousseaunian language and expressions, but rejected some of his views (Cornblit and Spector 1994, p. 35).

Manuel Belgrano (1770–1820) was the great champion of property in Argentina in the early nineteenth century. He was a lawyer and military hero educated at the universities of Salamanca and Valladolid. His reasoning and structure of thought owed considerably to the Late Scholastic works, which were still in vogue in Salamanca. Mariano Moreno (1778–1811), who translated and disseminated Rousseau, distanced himself from the Frenchman 'who had the disgrace of being delirious in his religious views'. Moreno did not translate the chapters on religion (ibid., p. 43).

Juan Bautista Alberdi (1810–84), the great champion of private property in the second half of the nineteenth century endorsed government support of religion because he found that:

> Only through it could one correct the defect that make the people incapable of liberty and government, such as: pride, an exaggerated sentiment of sufficiency,

> the susceptibility of its inhabitants, that does not allow them to admit and respect a truth that displeases them, coming from power or from liberty, be it heard by one citizen from another or as by a caretaker of power. This disposition makes political hatreds permanent, because the wounded pride has not learned to forget or distrust oneself. Without the domain over one self, without the authority of man over his will, which is the liberty of a citizen, discipline from one angle, authority can't exist … and without authority, the nation and society are chimerical notions. It is in religion that one finds the deep roots of amnesty, tolerance, abnegation and civil sacrifice. If half of the political order is found inside man, religion has the major part in the constitution of a country. (Alberdi 1895, p. 449)

Alberdi's ideas about the importance of private property, while they were implemented, helped Argentina become one of the most prosperous countries in the world. As religion is like a balsam that cures very slowly, Alberdi recommended that people should be exposed to it during childhood, and that one should start by the formation of an apostolate and the education of the clergy. In order to be more useful, it is in 'good seminaries rather than in splendid churches that government funds should be spent'. He favoured the introduction of foreign clergy, and he did not fear the Jesuits, which, according to him, exerted a beneficial impact in England and the United States. Although Alberdi did not recommend spending scarce government funds on religions that competed with Catholicism, he did recommend that the government, apart from granting them freedom, should use all other means to encourage them, from donation of public lands, to cemeteries, and other practical things (ibid., pp. 450–51).

In Fray Servando Teresa de Mier (1763–1827), Mexico also had an intellectual leader showing the compatibility between Christianity, natural law and private property (Cornblit and Spector 1994, p. 45). Fray Servando's many personal battles and exile, in addition to his overly candid critical stances, prevented him from becoming more influential during his time. When the state attacked the property rights of the Church, Lucas Alamán, much more relevant than Fray Servando in Mexican history, came to its strong defence (Alamán [1852] 1969, p. 540, cited by Cornblit and Spector 1994, p. 49).

Liberation theology and the social doctrine of the Church

From the late 1960s to the late 1980s, cultural and religious foundations of private property in Latin America suffered a direct attack by liberation theologians, and a weakening by the hierarchy. In every Latin American country, except perhaps Cuba, the Church is an influential institution. This influence is exerted mainly in the realm of faith. However, in their promotion of the religious message, the Church also exerts influence in the political and socio-economic fields. In 1968, the Latin American Bishops Confer-

ence met at Medellín, Colombia. The documents that came out of that meeting used a language that seemed hostile to private property and not so different from the one used by Pope Paul VI in his encyclical *Populorum Progressio* (1967).

Among the Latin Americans in the hierarchy, Msgr Alfonso López Trujillo of Colombia earned a reputation as a leading conservative bishop. As early as 1975, he wrote a book condemning some aspects of liberation theology. Unfortunately, at the time, López Trujillo's recipe for economic problems was the same one that has brought stagnation wherever it was applied. He argued that:

> [One] should encourage attempts, to renew forms for state, collective and social property which, technically and scientifically regulated, may bring remarkable advantages. In several instances, a certain process of nationalization, mainly of foreign corporations, may represent something healthy for the sovereignty and progress of our nations … There are many areas of production that up to now have been in the hands of certain persons or certain groups that clearly should become social or state property. (López Trujillo [1975] 1977, pp. 97–8)

López Trujillo scorned those who utter 'museum-like sounds such as "total freedom of trade"' and concluded by saying that 'we are convinced that capitalism is a human failure. We maintain some reservations about socialism which have nothing to do with the survival of neocapitalism that also crushes freedom, using another way to do it' (ibid., p. 101). Perhaps López Trujillo was also regarded as a conservative because he is also sceptical about centralized planning mechanisms. This scepticism seems to have been caused by the good influence of the writings of the eminent French Jesuit Pierre Bigo. Unfortunately, Bigo's influence was not strong enough. In 1976, López Trujillo stated that 'we cannot tolerate the widening gap in inequality and injustice. There has to be an advance in indispensable changes such as the bank and land-reforms'.[29]

López Trujillo's economic thought has been changing and his role in the Catholic Church continued to grow. His views are now totally compatible with those of the encyclical *Centesimus Annus*, where the Pope finds room for capitalism but, on anthropological grounds, disqualified socialism. Now a Cardinal, López Trujillo is, after the Pope, the person mostly responsible for having Professor Gary Becker, a Nobel laureate in economics, and a great champion of private property, become a member of the Pontifical Academy of Sciences.

The active role of the Vatican and the collapse of the Soviet empire, dealt a big blow to the radical attack on the institution of private property by liberation theologians. These contented themselves with weakening at private property, adopting the new creed of deep theology, health regulation and

redistribution, and attacking free trade. Free trade is nothing more than the right of people to freely exchange their private property across national borders.

Conclusion

We contributed to this volume because our knowledge of the history of religious ideas and our understanding of the historical aspects of Western civilization, made us aware of some of their connections with the institution of private property. We know how important it is to have a better understanding of the conditions needed for the establishment of private property, and the consistency between property and the dignity of the human being, a being whose spiritual nature can be glimpsed through Judeo-Christian traditions.

There are not yet enough studies, especially empirical, to show how countries with different religions and cultures respect private property. One of the most complete analyses, conducted by Luigi Zingales, Luigi Guiso and Paola Sapienza, showed that religious background had an impact on people's attitudes toward private property. They found that 'Christian religions are more positively associated with attitudes conducive to economic growth' (Zingales et al., 2003, p. 280). The results, however, were not what people expected. Catholics seemed to be the most respectful of private property. They also came on top in their defence of competition. Perhaps the problem is that apart from having a predisposition towards private property, individuals need to have attitudes that allow them to collaborate to build a rule of law, and it is in this regard that Catholics do not score well. They also scored very poorly on their understanding of the importance of incentives:

> On average, Christian religions are more positively associated with attitudes that are conducive to economic growth, while Islam is negatively associated. The ranking between the two main Christian denominations is less clear. Protestants trust others and the legal system more than Catholics and they are less willing to cheat on taxes and accept a bribe with respect to Catholics. By contrast, Catholics support private ownership twice as much as Protestants and are more in favour of competition than any other religious group (including Protestants). The only case in which Protestants seem more pro-market than Catholics is on incentives. When asked whether they are willing to accept more income inequality to provide incentives Protestants and Hindus are the only religious groups that favour incentives. (ibid., p. 280)

In another recent empirical study, results seemed the opposite. Countries with a predominant Catholic population 'protect the rights of creditors less than Protestant countries'. Religion had a higher predictive value for understanding which country respects more the private property rights of investors than international trade, language, income, or the origin of its legal system (Stultz and Williamson 2001).

The economic freedom indices also provide some data. They include respect of private property as one of the main elements, and show that the countries that have prospered the most are those that have been influenced by the notion of a rule of law as developed by Anglo-Saxon cultures. Yet, these indices are just over one decade old and need further refinement and testing.

Even when private property is recognized and protected by legal systems, the institution will depend on the lack of corruption in the implementation of the legal process. Although we have some indices of corruption, and have used them in previous studies, they are yet to have analytical and empirical strength. It is much easier to measure degrees of taxation than degrees of respect for the rule of law.

In recent decades, countries with a tradition based on common law have performed better than others. Nevertheless, from 1945 to 1979, Britain, whose people have almost unquestioned credentials for their role in bringing these traditions to different corners of the globe, had weakened its respect for private property in almost all imaginable forms. From nationalizations, to confiscatory taxation, British bureaucrats used many weapons in the arsenal against private property. From the prestige of its academic centres, the push to reduce the importance of private property was also exported to the countries in the Commonwealth and beyond. What was regarded as the cradle of liberty and the rule of law became useless. The so-called 'Thatcher revolution' slowed down this trend.

In the same way, while a couple of decades ago few Judeo-Christian leaders of moral authority could claim a role as defenders of property, the tide has also been turning on this front. The Catechism of the Catholic Church released during the last years of the twentieth century, states in point 2431 that:

> Economic activity, especially the activity of a market economy, cannot be conducted in an institutional, juridical, or political vacuum. On the contrary, it presupposes sure guarantees of individual freedom and private property, as well as a stable currency and efficient public services. Hence the principal task of the state is to guarantee this security, so that those who work and produce can enjoy the fruits of their labors and thus feel encouraged to work efficiently and honestly … Another task of the state is that of overseeing and directing the exercise of human rights in the economic sector. However, primary responsibility in this area belongs not to the state but to individuals and to the various groups and associations which make up society.

The institutions of private property have given human beings amazing opportunities to create wealth and enhance the human condition. Although many religions teach us to expect miracles, the future of private property will not depend so much on them but on how each person influences his/her culture and interprets its ties to their nature or Creator.

Notes

1. From the sketch of his life appearing in (Locke [1697] 1802).
2. A canon regular is a religious cleric destined to those works which relate to the devine mysteries, unlike monastic orders. In canon law, Jesuits, Theatines, and Oratorians come under canons regular concept as they are living by a rule but are not monks or friars.
3. *Clement of Alexandria: Quis dives salvatur?* (Who is the rich man that is being saved?) in Barnard (1897, pp. 1–66).
4. Ibid., pp. 28–9.
5. Lactantius, *Works* in (Roberts and Donaldson (eds), 1871).
6. Augustine of Hippo, *Works*, vol. XI, *Letters*, vol. III, *Letters* 156 and 157, in *The Fathers of the Church* (1951).
7. Salvian, *De Gubernatione Dei* (The government of God), *Writings of Salvian*, trans. Jeremiah O'Sullivan in *The Fathers of the Churc*h (1947).
8. *Burgundian Code* (1949). A *solidus* was the major gold coin established by Constantine the Great after the debasement of the coinage under Diocletian.
9. The *Burgundian Code* is a source for the property implications of the customary law punishments in Goebel (1976).
10. On this topic see, for example, Head and Landes (1987). The major developments in the legal history of medieval Western Europe can be found in the work of Berman (1983; French edn, 2002) and Merryman (1969). Lectures on legal subjects began in Bologna by the late eleventh century. The most significant of the lectures on canon law was presented about 1140 by Gratian as the *Harmony of Discordant Canons, or The Treatise on Laws* (Decretum DD. 1–20) (trans. Augustine Thompson O.P.) with the Ordinary gloss (trans. James Gordley) in *Studies in Medieval and Early Modern Canon Law* (1993). See also Radding (1988); Bellomo (1991, 1995).
11. Quotations from Bendix (1966, pp. 74, 417).
12. Refer to Collins (1986, ch. 3).
13. Carlyle and Carlyle (1950).
14. Compare Tuck (1979, p. 27); and Rothbard (1995, pp. 93–5).
15. See the rich recent literature on this subject such as Tuck (1979); Tierney (1982, 1997 and 1998); Gordley (1991); Rasmussen and Den Uyl (1991); Kuttner [1980] (1992); Brundage (1995); Miller (1995); Helmholz (1996); Noonan (1997); Reid (1998); Reid and Witte (1999).
16. Soto is speaking of the philosophers such as Plato and even Aristotle who favoured some sort of common ownership.
17. In Spanish it reads, '*se gasta lo que no se puede creer*'.
18. '*Imo praeceptum de non furando supponit rerum divisionem. Ergo rerum divisio non est contrarius naturale (alio quin ipso jure esset nulla) Quin potius approbata est in scripturis sacra*'. (Indeed, the commandment 'thou shall not steal' presupposes the division of things. The division of things, therefore, is not against natural law [if not, it would be, in itself illegal]. Otherwise, it would not be approved in the sacred scriptures.)
19. Ledesma (1614), Tratado VIII, Justicia Conmutativa.
20. '*Si inveniatur in loco, qui ad nullius particulare dominum pertineat, totus est inventoris*'.
21. '*Mineralia et venae auri, argenti et cuiusque metalli stando in iure naturae sunt domini fundi et in bonis ipsius*'.
22. '*Si ergo per voluntatem constituitur dominus, per eandem potest dominium ab se quodcunque abdicare*'.
23. Sir James Mackintosh, in 'Dissertation of the Progress of Ethical Philosophy, chiefly during the seventeenth and eighteenth centuries', in Mackintosh (1851, p. 23).
24. The Cambridge Historian Quentin Skinner has established the basis of modern natural rights thought to Late Medieval conciliarist canonists, especially Jean Gerson. Their thinking impacted on the early modern theorists, especially the School of Salamanca, culminating in the influential contributions of Francisco Suárez through Hugo Grotius. See Skinner (1978); also Brewer and Staves (1995). Hugo Grotius and Samuel Pufendorf

present a widely read understanding of natural rights and private property, see Tuck (1979).

25. Ashcroft (1990), Laslett (1957); reprinted in Yolton (1969); see Vaughn (1980). Locke was influenced by Francois Bernier, a fellow physician, who travelled to North Africa, the Near East and India, and wrote travel literature. Locke's extensive knowledge of travel literature informed his concepts of the state of nature and of private property. Locke's comparative approach continued during the eighteenth century including the studies of Adam Smith, Adam Ferguson and Lord Kames in the Scottish Enlightenment. See Schneider (1967); and Hamowy (1979) (1980). On Locke's study of travel literature, see Batz (1947); Ashcroft (1968, 1990, pp. 226–45); Harrison and Laslett (1965, pp. 1–61); Carus-Wilson (1954); North and Weingast (1989).
26. On this, see White (1978) and Gimpel (1976).
27. Balmés's *Fundamental Philosophy* was translated into English during the mid-nineteenth century and adopted as a textbook in many English-speaking seminaries.
28. Taparelli ([1855] 1928, point 979): '*Ma non basta a compiere la felicità civica abbondanza di ricchezze; quello che più monta è la tranquillità del possederle*'.
29. *El Tiempo*, Bogotá, 28 February 1976.

References

Acton, John Dalberg, (Lord) ([1907] 1993), *The History of Freedom and Other Essays*, New York: Classics of Liberty Library.

Alamán, Lucas ([1852] 1969), *Historia de Méjico* [History of Mexico], Méjico: Editorial Jus.

Alberdi, Juan Bautista (1895), 'Estudios económicos' [Economic Studies], in *Escritos Póstumos de J. B. Alberdi* [Posthumous writings of J.B. Alberdi], Buenos Aires: Imprenta Europea.

Albornóz, Bartolomé de (1573), *Arte de los Contratos* [Art of Contracts], Valencia.

Alcalá, Luis de (1543), *Tractado de los prestamos que passan entre mercaderes y tractantes, y por consiguiente de los logros, cambios, compras adelantadas, y ventas al fiado* [A treatise on the loans that are current between merchants and traders, and in consequence of the accomplishments, exchanges, advance purchases and sales installments.], Toledo: Juan de Ayala.

Antoine, P. Gabriel (1774), *Theologia Moralis Universa, 'De Iustitia et Iure'* [General Moral Theology], Krakow.

Aquinas, St Thomas ([1265–73] 1975), *Summa Theologica*, London: Blackfriars.

Aragón, Pedro de (1596), *De Iustitia et Iure* [On Justice and Law], Lyon.

Ashcroft, Richard (1968), 'Locke's State of Nature: historical fact or moral fiction', *American Political Science Review*, **62**, 898–915.

Ashcroft, Richard (1990), 'John Locke's library: portrait of an intellectual', in Jean S. Yolton (ed.), *A Locke Miscellany*, Bristol: Thoemmes.

Azpilcueta, Martin de (1556), *Manual de Confesores y Penitentes* [Manual of Confessors and Penitents], Salamanca.

Balmés, Jaime ([1844] 1925), 'Socialismo' [Socialism], in his *Obras Completas* [Complete Works], vol. XI, Barcelona: Biblioteca Balmes, pp. 265–6.

Barnard, P. Mordaunt (1897), *Clement of Alexandria: Quis dives salvetur*, Cambridge: Cambridge University Press.

Barnard, P. Mordaunt (ed.) ([1899] 1991), *The Biblical Text of Clement of Alexandria in the Four Gospels and the Acts of the Apostles*, Cambridge: Cambridge University Press.

Batz, William G. (1947), 'The historical anthropology of John Locke', *Journal of the History of Ideas*, **35**.

Bellomo, M. (1991), *L'Europa del Diritto Commune*, Roma: Il Cigno Galileo Galilei. *The Common Legal Past* (Ius Commune) *of Europe*, trans. L. Cochrane, Washington, DC: Catholic University of America Press, 1995.

Bendix, Richard (1966), *Max Weber: An Intellectual Portait*, London: University Paperbacks.

Berman, Harold J. (1983), *Law and Revolution: The Formation of the Western Legal Tradition*, Cambridge, MA: Harvard University Press. *Droit et Revolution*, French trans. Raoul Audouin, Aix-en-Provence: Librairie de l'Université d'Aix-en-Provence, 2002.

Blackstone, Sir William ([1765–69] 1962), *Commentaries on the Laws of England*, Boston, MA: Beacon Press.
Brewer, John and Susan Staves (eds) (1995), *Early Modern Concepts of Property*, London: Routledge.
Brundage, James A. (1995), *Medieval Canon Law*, London, New York: Longman.
Burgundian Code (1949) (trans. Katherine Fischer Drew), Philadelphia, PA: University of Pennsylvania Press.
Carlyle, Sir R.W. and A.J. Carlyle (1950), *The Political Theory of the Thirteenth Century*, vol. V of *A History of Medieval Political Theory in the West*, Edinburgh and London: William Blackwood & Sons, pp. 14–20.
Carus-Wilson, E.M. (ed.), (1954), *Essays in Economic History*, vol. I, London: E. Arnold.
Collins, Randall (1986), 'The Weberian revolution of the High Middle Ages', ch. 3, in *Weberian Sociological Theory*, Cambridge: Cambridge University Press.
Cornblit, Oscar and Horacio Spector (1994), *Fundamentos de la Propiedad* [Foundations of Property], Buenos Aires: CEPPA.
Courcelle-Seneuil, J.G. (1887), *Préparation a L'étude du droit*, Paris: Guillaumin.
Encyclical (1931), *Quadragesimo Anno*, 120, 15 May.
Escobar y Mendoza, Antonio de (1662), *Universae Theologiae Moralis, Tomi Quinti pars prima, Iustitia et Iure* [General Moral Theology, 5th book, first part on Justice and Law], Lyon.
Fustel de Coulanges, Numa Denis ([1864] 1956), *The Ancient City: A Study, of the Religion, Laws, and Institutions of Greece and Rome*, Garden City, NY: Doubleday Anchor Books.
García, Francisco (1583), *Tratado Utilísimo de Todos los Contratos, Quantos en los Negocios Humanos se Pueden Ofrecer* [Very useful Treatise on all the Contracts that might be brought up in Human Business], Valencia.
Gerson, Jean (1962), *Oeuvres Completes , Vol. 3, L'Oeuvre Magistrale (87–105)*, Introduction, texte et notes por Mgr Glorieux, Tournai, Paris: Desclée et Cie.
Gimpel, Jean (1976), *The Medieval Machine: The Industrial Revolution of the Middle Ages*, New York: Holt, Rinehart & Winston.
Goebel, Jr., Julius [1937] 1976), *Felony and Misdemeanor*, Philadelphia: University of Pennsylvania Press.
Goguet, Antoine Yves ([1758], [1775] reprint c.1976), *The Origins of Laws, Arts, and Sciences, and their Progress Among the Most Ancient Nations*, New York: AMS Press.
Gordley, James (1991), *The Philosophical Origins of Modern Contract Doctrine*, Oxford: Clarendon.
Gratian, 12th cent. (1993), *Studies in Medieval and Early Modern Canon Law*, trans. Augustine Thompson, Washington, DC: Catholic University of America Press.
Grice-Hutchinson, Marjorie (1978), *Early Economic Thought in Spain*, London: George Allen & Unwin.
Hamowy, R. (1979), 'Jefferson and the Scottish Enlightenment', *William and Mary Quarterly*, **36**, pp. 503–23.
Hamowy, R. (1980), 'Communications', *William and Mary Quarterly*, **37**, pp. 534–40.
Harrison John and Peter Laslett (eds) (1965), *The Library of John Locke*, Oxford: Oxford Bibliographical Society Publications, New Series, pp. 1–61.
Head, Thomas and Richard Landes (1987), 'Essays on the Peace of God: The Church and the People in Eleventh-Century France', in *Historical Reflections/Reflexions Historiques*, Waterloo, Ontario: vol. 14, No. 3.
Helmholz, Richard R. (1996), *The Spirit of the Classical Canon Law*, Athens, GA: University of Georgia Press.
Ireland, John (1905), *The Church and Modern Society Lectures and Addresses by John Ireland*, Minneapolis-St Paul, MN: Pioneer Press.
Jarrett, Bede (1942), *Social Theories of the Middle Ages: 1200–1500*, Westminster, MD: Newman Book Shop.
John of Paris ([c. 1302] 1974), *On Royal and Papal Power*, New York: Columbia University Press.

Kuttner, Stephan ([1980] 1992), *The History of Ideas and Doctrines of Canon Law in the Middle Ages* (2nd edn) Aldershot, Hampshire: Variorum Reprints.
La Cruz, Felipe de (1637), *Tratado Unico de Intereses Sobre si se Puede Llevar Dinero por Prestallo* [A Single Treatise on Interest and if one can ask money for Lending it], Madrid: Francisco Martinez.
Ladner, Gerhart B. (1959), *The Idea of Reform: Its Impact on Christian Thought and Action in the Age of the Fathers* , Cambridge, MA: Harvard University Press.
Lal, Deepak (1998), *Unintended Consequences*: *The Impact of Factor Endowments, Culture, and Politics on Long-Run Economic Performance*, Cambridge, MA: MIT Press.
Laslett, Peter (1957), 'John Locke, the Great Recoinage, and the origins of the Board of Trade: 1695–1698', *William and Mary Quarterly*, **14**, 370–92.
Ledesma, Pedro de (1614), *Summa*, Salamanca.
Lessio, Leonardo (1605), *De Iustitia et Iure* [On Justice and Law], Louvain: Masy.
Liberatore, Matteo (1891), *Principles of Political Economy*, New York: Benzinger.
Locke, John ([1690] 1970), *Two Treatises of Government*, ed. Peter Laslett, 2nd edn, Cambridge: Cambridge University Press, II.
Locke, John ([1690] 1946), *The Second Treatise of Civil Government and A Letter Concerning Toleration*, ed. J.W. Gooch, Oxford: Blackwell.
Locke, John **(**[1695] 1989), *The Reasonableness of Christianity*, Washington, DC: Gateway Editions.
Locke, John ([1697] 1802), *The Conduct of the Understanding*, London: M. Jones.
López Trujillo, Alfonso ([1975] 1977), *Liberation or Revolution, An Examination of the Priest's Role in the Socioeconomic Class Struggle in Latin America*, Huntington, Indiana: Our Sunday Visitor, Inc.
Macfarlane, Alan ([1978] 1979), *The Origins of English Individualism*, New York: Cambridge University Press.
Mackintosh, James (1851), *The Miscellaneous Works of Sir James Mackintosh*, London: Longman, Brown, Green & Longmans.
Mariana, Juan de ([c. 1600] 1950), *Discurso de las Cosas de la Compañia* [A Discourse on the Affairs of the Company of Jesus], Biblioteca de Autores Españoles, Madrid: Atlas.
Melitz, Jacques and Donald Winch (eds) (1978), *Religious Thought and Economic Society*, Durham, NC: Duke University Press.
Mercado, Tomás de ([1569] 1571), *Summa de Tratos y Contratos* [Summa of Contracts and Trades], Seville.
Merryman, John Henry (1969), *The Civil Law Tradition*, Stanford, CA: Stanford University Press.
Miller, Jr., Fred D. (1995), *Nature, Justice, and Rights in Aristotle's Politics*, Oxford: Clarendon.
Molina, Luis de ([1593] 1614), *De Iustitia et Iure* [On Justice and Law], Moguntiae.
Nell-Breuning, Oswald von (1937), *Reorganization of Social Economy. The Social Encyclical Developed and Explained*, translated and edited by Bernard Dempsey, S.J., Milwaukee, WI: Herder.
Nisbet, Robert (1979), 'The idea of progress', *Literature of Liberty*, **2** (1), 7–37.
Nisbet, Robert (1980), *History of the Idea of Progress*, New York: Basic Books.
Noonan, John T. (1997), *Canons and Canonists in Context*, Goldbach: Keip.
North, Douglass 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** (4), December, reprinted in Paul W. Brake and Mathew D. McCubbins (eds) (1998), *The Origins of Liberty: Political and Economic Liberalization of the Modern World*, ch. 3, Princeton, NJ: Princeton University Press, pp. 16–47.
Noyes, C. Reinold (1936), *The Institution of Property*, New York: Longmans, Green & Co.
Rasmussen, Douglass and Douglass Den Uyl (1991), *Liberty and Nature: An Aristotelian Defense of Liberal Order*, La Salle, IL: Open Court.
Radding, Charles M. (1988), *The Origins of Medieval Jurisprudence: Pavia and Bologna 850–1150*, New Haven, CT: Yale University Press.
Reid, Jr., Charles (1998), 'The medieval origins of the Western natural rights tradition: the achievement of Brian Tierney,' *Cornell Law Review*, **83**, 437.

Reid, Jr., Charles J. and J. Witte, Jr. (1999), 'In the steps of Gratian: writing the history of canon law in the 1990s', *Emory Law Journal*, **48** (2), pp. 647–88.
Roberts, Alexander and James Donaldson (eds) (1871), *Ante-Nicene Fathers*, Edinburgh: T. & T. Clark.
Rosmini, Antonio ([1845] 1967), *Filosofia del diritto* [Philosophy of Law], ed. R. Orechia, Padua: Cedam.
Rothbard, Murray Newton (1995), *Economic Thought Before Adam Smith*, Aldershot, UK and Brookfield, US VT: Edward Elgar.
Ryan, Alan (1984), *Property and Political Theory*, Oxford: Basil Blackwell.
Salón, Miguel ([1581] 1591), *Commentariorum in Disputationem de Iustitia Quam Habet D. Tho. Secunda Sectione Secundae parties suae Summa Theologicae* [A commentary on St. Thomas Aquinas' investigations on Justice as contained in the second part of the second section of his Summa Theologica], Valencia.
Saravia de la Calle, Luis (Veronense) ([1544] 1949), *Instrucción de Mercaderes muy Provechosa* [Very Fruitful Instruction for Merchants], Medina del Campo: Pedro Castro.
Schneider, Louis (ed.) (1967), *The Scottish Moralists on Human Nature and Society*, Chicago: University of Chicago Press.
Schumpeter, Joseph (1954), *History of Economic Analysis*, Oxford and New York: Oxford University Press.
Skinner, Quentin (1978), *The Foundations of Modern Political Thought*, 2 vols, Cambridge: Cambridge University Press.
Soto, Domingo de ([1553] 1968), *De Iustitia et Iure* [On Justice and Law], Madrid: IEP.
Stultz, René M. and R. Williamson (2001), 'Culture, openness, and finance', NBER working paper 8222, National Bureau of Economic Research.
Sumner Maine, Sir Henry ([1861] 1986), *Ancient Law*, New York: Dorset Press.
Taparelli, Luigi ([1855] 1928), *Saggio Teoretico di Dritto Naturale* [Theoretical Study of Natural Law], Rome: Civilta Cattolica.
The Fathers of the Church (1947, 1951), New York: CIMA Publishing.
Tierney, Brian ([1955] 1998), *Foundations of the Conciliar Theory: The Contribution of the Medieval Canonists from Gratian to the Great Schism*, Leiden, New York: Brill.
Tierney, Brian (1982), *Religion, Law, and the Growth of Constitutional Thought, 1150–1650*, Cambridge, New York: Cambridge University Press.
Tierney, Brian (1997), *The Idea of Natural Rights: Studies in Natural Right, Natural Law, and Church Law, 1150–1625*, Atlanta, GA: Scholar Press.
Tuck, Richard (1979), *Natural Rights Theories: Their Origin and Development*, Cambridge: Cambridge University Press.
Vaughn, Karen (1980), *John Locke: Economist and Social Scientist*, Chicago, IL: University of Chicago Press.
Villalobos, Henrique de (1632), *Summa de la Theologia Moral y Canónica* [A Summa of Moral and Canonical Theology], Barcelona.
Villalón, Cristóbal de (1542), *Provechoso Tratado de Cambios y Contrataciones de Mercaderes y Reprobacion de Usura* [A Useful Treatise of Merchant Contracts and Exchanges and a reprobation of usury], Valladolid: Francisco Fernández de Córdoba.
Vitoria, Francisco de ([c. 1530] 1934), *De Justitia* [On Justice], ed. Beltrán de Heredia, Madrid: Publicaciones de la Asociación Francisco de Vitoria.
White, Jr., Lynn (1978), *Medieval Religion and Technology*, Berkeley, CA: University of California Press.
Yolton, Jean (ed.) (1969), *John Locke: Problems and Perspectives*, Cambridge: Cambridge University Press.
Zingales, L., L. Guiso and P. Sapienza (2003) 'People's opium? Religion and economic attitudes', *Journal of Monetary Economics*, **50**, 225–82.

2 The ethics and economics of private property

Hans-Hermann Hoppe

The problem of social order

Alone on his island, Robinson Crusoe can do whatever he pleases. For him, the question concerning rules of orderly human conduct – social cooperation – simply does not arise. Naturally, this question can only arise once a second person, Friday, arrives on the island. Yet even then, the question remains largely irrelevant so long as no scarcity exists. Suppose the island is the Garden of Eden; all external goods are available in superabundance. They are 'free goods', just as the air that we breathe is normally a 'free' good. Whatever Crusoe does with these goods, his actions have repercussions neither with respect to his own future supply of such goods nor regarding the present or future supply of the same goods for Friday (and vice versa). Hence, it is impossible that there could ever be a conflict between Crusoe and Friday concerning the use of such goods. A conflict is only possible if goods are scarce. Only then will there arise the need to formulate rules that make orderly – conflict-free – social cooperation possible.

In the Garden of Eden only two scarce goods exist: the physical body of a person and its standing room. Crusoe and Friday each have only one body and can stand only at one place at a time. Hence, even in the Garden of Eden conflicts between Crusoe and Friday can arise: Crusoe and Friday cannot occupy the same standing room simultaneously without coming thereby into physical conflict with each other. Accordingly, even in the Garden of Eden rules of orderly social conduct must exist – rules regarding the proper location and movement of human bodies. And outside the Garden of Eden, in the realm of scarcity, there must be rules that regulate not only the use of personal bodies but also of everything scarce so that all possible conflicts can be ruled out. This is the problem of social order.

The solution: private property and original appropriation

In the history of social and political thought, various proposals have been advanced as a solution to the problem of social order, and this variety of mutually inconsistent proposals has contributed to the fact that today's search for a single 'correct' solution is frequently deemed illusory. Yet as I shall try to demonstrate, a correct solution exists; hence, there is no reason to succumb to moral relativism. The solution has been known for hundreds of years, if

not much longer.[1] In modern times this old and simple solution was formulated most clearly and convincingly by Murray N. Rothbard.[2]

Let me begin by formulating the solution – first for the special case represented by the Garden of Eden and subsequently for the general case represented by the 'real' world of all-around scarcity – and then proceed to the explanation of why this solution, and no other, is correct.

In the Garden of Eden, the solution is provided by the simple rule stipulating that everyone may place or move his own body wherever he pleases, provided only that no one else is already standing there and occupying the same space. And outside of the Garden of Eden, in the realm of all-around scarcity the solution is provided by this rule: everyone is the proper owner of his own physical body as well as of all places and nature-given goods that he occupies and puts to use by means of his body, provided that no one else has already occupied or used the same places and goods before him. This ownership of 'originally appropriated' places and goods by a person implies his right to use and transform these places and goods in any way he sees fit, provided that he does not thereby forcibly change the physical integrity of places and goods originally appropriated by another person. In particular, once a place or good has been first appropriated, in John Locke's words, by 'mixing one's labor' with it, ownership in such places and goods can be acquired only by means of a voluntary – contractual – transfer of its property title from a previous to a later owner.

In light of widespread moral relativism, it is worth pointing out that this idea of original appropriation and private property as a solution to the problem of social order is in complete accordance with our moral 'intuition'. Is it not simply absurd to claim that a person should not be the proper owner of his body and the places and goods that he originally, that is, prior to anyone else, appropriates, uses and/or produces by means of his body? For who else, if not he, should be their owner? And is it not also obvious that the overwhelming majority of people – including children and primitives – in fact act according to these rules, and do so as a matter of course?

Moral intuition, as important as it is, is not proof. However, there also exists proof of the veracity of our moral intuition.

The proof is twofold. On the one hand, the consequences that follow if one were to deny the validity of the institution of original appropriation and private property are spelled out: if person A were *not* the owner of his own body and the places and goods originally appropriated and/or produced with this body as well as of the goods voluntarily (contractually) acquired from another previous owner, then only two alternatives would exist. Either *another* person, B, must be recognized as the owner of A's body as well as the places and goods appropriated, produced or acquired by A, or both persons, A and B, must be considered equal co-owners of all bodies, places and goods.

In the first case, A would be reduced to the rank of B's slave and object of exploitation. B would be the owner of A's body and all places and goods appropriated, produced and acquired by A, but A in turn would not be the owner of B's body and the places and goods appropriated, produced and acquired by B. Hence, under this ruling two categorically distinct classes of persons would be constituted – *Untermenschen* such as A and *Übermenschen* such as B – to whom different 'laws' apply. Accordingly, such a ruling must be discarded as a human ethic equally applicable to everyone *qua* human being (rational animal). From the very outset, any such ruling is recognized as not universally acceptable and thus cannot claim to represent law. For a rule to aspire to the rank of a law – a 'just' rule – it is necessary that such a rule apply equally and universally to everyone.

Alternatively, in the second case of universal and equal co-ownership, the requirement of equal law for everyone would be fulfilled. However, this alternative would suffer from an even more severe deficiency, because if it were applied, all of mankind would instantly perish. (Since every human ethic must permit the survival of mankind, this alternative must also be rejected.) Every action of a person requires the use of some scarce means (at least of the person's body and its standing room), but if all goods were co-owned by everyone, then no one, at no time and no place, would be allowed to do anything unless he had previously secured every other co-owner's consent to do so. Yet how could anyone grant such consent were he not the exclusive owner of his own body (including his vocal chords) by which means his consent must be expressed? Indeed, he would first need another's consent in order to be allowed to express his own, but these others could not give their consent without first having his, and so it would go on.

This insight into the praxeological impossibility of 'universal communism', as Rothbard referred to this proposal, brings me immediately to an alternative way of demonstrating the idea of original appropriation and private property as the only correct solution to the problem of social order.[3] Whether or not people have any rights and, if so, which ones, can only be decided in the course of argumentation (propositional exchange). Justification – proof, conjecture, refutation – is argumentative justification. Anyone who denied this proposition would become involved in a performative contradiction because his denial would itself constitute an argument. Even an ethical relativist would have to accept this first proposition, which is referred to accordingly as the a priori of argumentation.

From the undeniable acceptance – the axiomatic status – of this a priori of argumentation, two equally necessary conclusions follow. First, it follows from the a priori of argumentation when there is no rational solution to the problem of conflict arising from the existence of scarcity. Suppose in my earlier scenario of Crusoe and Friday that Friday were not the name of a man

but of a gorilla. Obviously, just as Crusoe could face conflict regarding his body and its standing room with Friday the man, so might he with Friday the gorilla. The gorilla might want to occupy the same space that Crusoe already occupied. In this case, at least if the gorilla were the sort of entity that we know gorillas to be, there would be no rational solution to their conflict. Either the gorilla would push aside, crush, or devour Crusoe – that would be the gorilla's solution to the problem – or Crusoe would tame, chase, beat, or kill the gorilla – that would be Crusoe's solution. In this situation, one might indeed speak of moral relativism. However, it would be more appropriate to refer to this situation as one in which the question of justice and rationality simply would not arise; that is, it would be considered an extra-moral situation. The existence of Friday the gorilla would pose a technical, not a moral, problem for Crusoe. He would have no other choice than to learn how to successfully manage and control the movements of the gorilla just as he would have to learn to manage and control other inanimate objects of his environment.

By implication, only if both parties in a conflict are capable of engaging in argumentation with one another can one speak of a moral problem, and is the question of whether or not there exists a solution to it a meaningful question. Only if Friday, regardless of his physical appearance, is capable of argumentation (even if he has shown himself to be capable only once), can he be deemed rational and does the question whether or not a correct solution to the problem of social order exists make sense. No one can be expected to give any answer to someone who has never raised a question or, more to the point, who has never stated his own relativistic viewpoint in the form of an argument. In that case, this 'other' cannot but be regarded and treated as an animal or plant, that is, as an extra-moral entity. Only if this other entity can pause in his activity, whatever it might be, step back, and say 'yes' or 'no' to something one has said, do we owe this entity an answer and, accordingly, can we possibly claim that our answer is the correct one for both parties involved in a conflict.

Moreover, it follows from the a priori of argumentation that everything that must be presupposed in the course of an argumentation as the logical and praxeological precondition of argumentation cannot in turn be argumentatively disputed as regards its validity without the arguer becoming thereby entangled in an internal (performative) contradiction.

Now, propositional exchanges are not made up of free-floating propositions, but rather constitute a specific human activity. Argumentation between Crusoe and Friday requires that both have, and mutually recognize each other as having, exclusive control over their respective bodies (their brain, vocal chords and so on) as well as the standing room occupied by their bodies. No one could propose anything and expect the other party to con-

vince himself of the validity of this proposition or deny it and propose something else unless his and his opponent's right to exclusive control over their respective bodies and standing rooms were presupposed. In fact, it is precisely this mutual recognition of the proponent's as well as the opponent's property in his own body and standing room which constitutes the *characteristicum specificum* of all propositional disputes: that while one may not agree regarding the validity of a specific proposition, one can agree nonetheless on the fact that one disagrees. Moreover, this right to property in one's own body and its standing room must be considered a priori (or indisputably) justified by proponent and opponent alike. Anyone who claimed any proposition as valid *vis-à-vis* an opponent would already presuppose his and his opponent's exclusive control over their respective body and standing room simply in order to say 'I claim such and such to be true, and I challenge you to prove me wrong'.

It would be equally impossible to engage in argumentation and rely on the propositional force of one's arguments if one were not allowed to own (exclusively control) other scarce means (besides one's body and its standing room). If one did not have such a right, then we would all immediately perish and the problem of justifying rules – as well as any other human problem – would simply not exist. Hence, by virtue of the fact of being alive property rights to other things must be presupposed as valid, too. No one who is alive can possibly argue otherwise.

Furthermore, if a person were not permitted to acquire property in these goods and spaces by means of an act of original appropriation, that is, by establishing an objective (intersubjectively ascertainable) link between himself and a particular good and/or space prior to anyone else, and if instead property in such goods or spaces were granted to latecomers, then no one would ever be permitted to begin using any good unless he had previously secured such a latecomer's consent. Yet how can a latecomer consent to the actions of an early comer? Moreover, every latecomer would in turn need the consent of other and later latecomers, and so on. That is, we, our forefathers, or our progeny would not have been or would not be able to survive if one followed this rule. However, in order for any person – past, present or future – to argue anything, survival must be possible; and in order to do just this property rights cannot be conceived of as being timeless and unspecific with respect to the number of persons concerned. Rather, property rights must necessarily be conceived of as originating by means of action at definite points in time and space by definite individuals. Otherwise, it would be impossible for anyone to ever say anything at a definite point in time and space and for someone else to be able to reply. Simply saying, then, that the first-user–first-owner rule of the ethics of private property can be ignored or is unjustified implies a performative contradiction, as one's being able to say

so must presuppose one's existence as an independent decision-making unit at a given point in time and space.[4]

Misconceptions and clarifications

According to this understanding of private property, property ownership means the exclusive control by a particular person over specific physical objects and spaces. Conversely, property rights invasion means the uninvited physical damage or diminution of things and territories owned by other persons. In contrast, a widely held view holds that the damage or diminution of the value (or price) of someone's property also constitutes a punishable offense.

As far as the (in)compatibility of both positions is concerned, it is easy to recognize that nearly every action of an individual can alter the value (price) of someone else's property. For example, when person A enters the labor or the marriage market, this may change the value of B in these markets. And when A changes his relative valuations of beer and bread, or if A himself decides to become a brewer or baker, this changes the value of the property of other brewers and bakers. According to the view that value damage constitutes a rights violation, A would be committing a punishable offense *vis-à-vis* brewers or bakers. If A is guilty, then B and the brewers and bakers must have the right to defend themselves against A's actions, and their defensive actions can only consist of physical invasions of A and his property. B must be permitted to physically prohibit A from entering the labor or marriage market; the brewers and bakers must be permitted to physically prevent A from spending his money as he sees fit. However, in this case the physical damage or diminution of the property of others cannot be viewed as a punishable offense. Since physical invasion and diminution are defensive actions, they are legitimate. Conversely, if physical damage and diminution constitute a rights violation, then B or the brewers and bakers do not have the right to defend themselves against A's actions, for his actions – his entering of the labor and marriage market, his altered evaluation of beer and bread, or his opening of a brewery or bakery – do not affect B's bodily integrity or the physical integrity of the property of brewers or bakers. If they physically defend themselves nonetheless, then the right to defense would lie with A. In that case, however, it cannot be regarded as a punishable offense if one alters the value of other people's property. A third possibility does not exist.

Both ideas of property rights are not only incompatible, however. The alternative view – that one could be the owner of the value or price of scarce goods – is indefensible. While a person has control over whether or not his actions will change the physical properties of another's property, he has no control over whether or not his actions affect the value (or price) of another's property. This is determined by other individuals and their evaluations. Consequently, it would be impossible to know in advance whether or not one's

planned actions were legitimate. The entire population would have to be interrogated to ensure that one's actions would not damage the value of someone else's property, and one could not begin to act until a universal consensus had been reached. Mankind would die out long before this assumption could ever be fulfilled.

Moreover, the assertion that one has a property right in the value of things involves a contradiction, for in order to claim this proposition to be valid – universally agreeable – it would have to be assumed that it is permissible to act before agreement is reached. Otherwise, it would be impossible to ever propose anything. However, if one is permitted to assert a proposition – and no one could deny this without running into contradictions – then this is only possible because physical property borders exist, that is, borders which everyone can recognize and ascertain independently and in complete ignorance of others' subjective valuations.[5]

Another, equally common misunderstanding of the idea of private property concerns the classification of actions as permissible or impermissible based exclusively on their physical effects, that is, without taking into account that every property right has a history (temporal genesis).

If A currently physically damages the property of B (for example, by air pollution or noise), the situation must be judged differently depending on whose property right was established earlier. If A's property was founded first, and if he had performed the questionable activities before the neighboring property of B was founded, then A may continue with his activities. A has established an easement. From the outset, B had acquired dirty or loud property, and if B wants to have his property clean and quiet he must pay A for this advantage. Conversely, if B's property was founded first, then A must stop his activities; and if he does not want to do this, he must pay B for this advantage. Any other ruling is impossible and indefensible because as long as a person is alive and awake, he cannot not act. An early comer cannot, even if he wished otherwise, wait for a latecomer and his agreement before he begins acting. He must be permitted to act immediately. And if no other property besides one's own exists (because a latecomer has not yet arrived), then one's range of action can be deemed limited only by laws of nature. A latecomer can only challenge the legitimacy of an early comer if he is the owner of the goods affected by the early comer's actions. However, this implies that one can be the owner of unappropriated things; that is, that one can be the owner of things one has not yet discovered or appropriated through physical action. This means that no one is permitted to become the first user of a previously undiscovered and unappropriated physical entity.

The economics of private property

The idea of private property not only agrees with our moral intuitions and is the sole just solution to the problem of social order; the institution of private property is also the basis of economic prosperity and of 'social welfare'. As long as people act in accordance with the rules underlying the institution of private property, social welfare is optimized.

Every act of original appropriation improves the welfare of the appropriator (at least *ex ante*); otherwise, it would not be performed. At the same time, no one is made worse off by this act. Any other individual could have appropriated the same goods and territories if only he had recognized them as scarce, and hence, valuable. However, since no other individual made such an appropriation, no one else can have suffered a welfare loss on account of the original appropriation. Hence, the so-called 'Pareto criterion' (that it is scientifically legitimate to speak of an improvement of 'social welfare' only if a particular change increases the individual welfare of at least one person and leaves no one else worse off) is fulfilled. An act of original appropriation meets this requirement. It enhances the welfare of one person, the appropriator, without diminishing anyone else's physical wealth (property). Everyone else has the same quantity of property as before and the appropriator has gained new, previously non-existent property. In so far, an act of original appropriation always increases social welfare.

Any further action with originally appropriated goods and territories enhances social welfare, for no matter what a person does with his property, it is done to increase his welfare. This is the case when he consumes his property as well as when he produces new property out of 'nature'. Every act of production is motivated by the producer's desire to transform a less valuable entity into a more valuable one. As long as acts of consumption and production do not lead to the physical damage or diminution of property owned by others, they are regarded as enhancing social welfare.

Finally, every voluntary exchange (transfer) of appropriated or produced property from one owner to another increases social welfare. An exchange of property is only possible if both owners prefer what they acquire over what they surrender and thus expect to benefit from the exchange. Two persons gain in welfare from every exchange of property, and the property under the control of everyone else is unchanged.

In distinct contrast, any deviation from the institution of private property must lead to social welfare losses.

In the case of universal and equal co-ownership – universal communism instead of private property – the price to be paid would be mankind's instant death because universal co-ownership would mean that no one would be allowed to do anything or move anywhere. Each actual deviation from a private property order would represent a system of unequal domination and

hegemony. That is, it would be an order in which one person or group – the rulers, exploiters or *Übermenschen* – would be permitted to acquire property other than by original appropriation, production or exchange, while another person or group – the ruled, exploited or *Untermenschen* – would be prohibited from doing likewise. While hegemony is possible, it would involve social welfare losses and would lead to relative impoverishment.

If A is permitted to acquire a good or territory which B has appropriated as indicated by visible signs, the welfare of A is increased at the expense of a corresponding welfare loss on the part of B. The Pareto criterion is not fulfilled, and social welfare is suboptimal. The same is true with other forms of hegemonic rule. If A prohibits B from originally appropriating a hitherto unowned piece of nature; if A may acquire goods produced by B without B's consent; if A may proscribe what B is permitted to do with his appropriated or produced goods (apart from the requirement that one is not permitted to physically damage or diminish others' property) – in each case there is a 'winner', A, and a 'loser', B. In every case, A increases his supply of property at the expense of B's corresponding loss of property. In no case is the Pareto criterion fulfilled, and a suboptimal level of social welfare always results.

Moreover, hegemony and exploitation lead to a reduced level of future production. Every ruling which grants non-appropriators, non-producers and non-traders control, either partial or full, over appropriated, produced or traded goods, leads necessarily to a reduction of future acts of original appropriation, production and mutually beneficial trade. For the person performing them, each of these activities is associated with certain costs, and the cost of performing them increases under a hegemonic system and that of not performing them decreases. Present consumption and leisure become more attractive as compared to production (future consumption), and the level of production will fall below what it otherwise would have been. As for the rulers, the fact that they can increase their wealth by expropriating property appropriated, produced or contractually acquired by others will lead to a wasteful usage of the property at its disposal. Because they are permitted to supplement their future wealth by means of expropriation (taxes), present-orientation and consumption (high time preference) is encouraged, and insofar as they use their goods 'productively' at all, the likelihood of misallocations, miscalculation and economic loss is systematically increased.

The classic pedigree

As noted at the outset, the ethics and economics of private property presented above does not claim originality. Rather, it is a modern expression of a 'classic' tradition, going back to its beginnings in Aristotle, Roman law, Aquinas, the late Spanish Scholastics, Grotius and Locke.[6]

In contrast to the communist utopia of Plato's *Republic*, Aristotle provides a comprehensive list of the comparative advantages of private property in *Politics*. First, private property is more productive:

> What is common to the greatest number gets the least amount of care. Men pay most attention to what is their own; they care less for what is common; or at any rate they care for it only to the extent to which each is individually concerned. Even when there is no other cause for inattention, men are more prone to neglect their duty when they think that another is attending to it.[7]

Second, private property prevents conflict and promotes peace. When people have their own separate domains of interest, 'there will not be the same grounds for quarrels, and the amount of interest will increase, because each man will feel that he is applying himself to what is his'.[8] 'Indeed, it is a fact of observation that those who own common property, and share in its management, are far more often at variance with one another than those who have property in severalty'.[9] Further, private property has existed always and everywhere, whereas nowhere have communist utopias sprung up spontaneously. Finally, private property promotes the virtues of benevolence and generosity. It allows one to be so with friends in need.

Roman law, from the *Twelve Tables* to the *Theodosian Code* and the *Justinian Corpus*, recognized the right of private property as near absolute. Property stemmed from unchallenged possession, prior usage established easements, a property owner could do with his property as he saw fit, and freedom of contract was acknowledged. As well, Roman law distinguished importantly between 'national' (Roman) law – *ius civile* – and 'international' law – *ius gentium*.

The Christian contribution to this classic tradition – embodied in St Thomas Aquinas and the late Spanish Scholastics as well as Protestants Hugo Grotius and John Locke – is twofold. Both Greece and Rome were slave-holding civilizations. Aristotle, characteristically, considered slavery a natural institution. In contrast, Western – Christian – civilization, notwithstanding some exceptions, has been essentially a society of free men. Correspondingly, for Aquinas as for Locke, every person had a proprietary right over himself (self-ownership). Moreover, Aristotle, and classic civilization generally, were disdainful of labor, trade and money-making. In contrast, in accordance with the Old Testament, the Church extolled the virtues of labor and work. Correspondingly, for Aquinas as for Locke, it was by work, use and cultivation of previously unused land that property first came into existence.

This classic theory of private property, based on self-ownership, original appropriation (homesteading), and contract (title transfer), continued to find prominent proponents, such as J.B. Say. However, from the height of its

influence in the eighteenth century until quite recently, with the advance of the Rothbardian movement, the classic theory had slipped into oblivion.

For two centuries, economics and ethics (political philosophy) had diverged from their common origin in natural law doctrine into seemingly unrelated intellectual endeavors. Economics was a value-free 'positive' science. It asked 'what means are appropriate to bring about a given (assumed) end?'. Ethics was a 'normative' science (if it was a science at all). It asked 'what ends (and what use of means) is one justified to choose?'. As a result of this separation, the concept of property increasingly disappeared from both disciplines. For economists, property sounded too normative; for political philosophers property smacked of mundane economics.

In contrast, Rothbard noted, such elementary economic terms as direct and indirect exchange, markets and market prices as well as aggression, crime, tort and fraud cannot be defined or understood without a theory of property. Nor is it possible to establish the familiar economic theorems relating to these phenomena without the implied notion of property and property rights. A definition and theory of property must precede the definition and establishment of all other economic terms and theorems.

Rothbard's unique contribution, from the early 1960s until his death in 1995, was the rediscovery of property and property rights as the common foundation of both economics and political philosophy, and the systematic reconstruction and conceptual integration of modern, marginalist economics and natural-law political philosophy into a unified moral science: libertarianism.

Chicago diversions

At the time when Rothbard was restoring the concept of private property to its central position in economics and reintegrating economics with ethics, other economists and legal theorists associated with the University of Chicago such as Ronald Coase, Harold Demsetz and Richard Posner were also beginning to redirect professional attention to the subject of property and property rights.[10]

However, whereas for Rothbard private property and ethics logically precede economics, for the latter private property and ethics are subordinate to economics and economic considerations. According to Posner, whatever increases social wealth is just.[11]

The difference between the two approaches can be illustrated considering one of Coase's problem cases: a railroad runs beside a farm; the engine emits sparks, damaging the farmer's crop; what is to be done?

From the classic viewpoint, what needs to be established is who was there first, the farmer or the railroad? If the farmer was there first, he could force the railroad to cease and desist or demand compensation. If the railroad was

there first, then it might continue emitting sparks and the farmer would have to pay the railroad to be spark free.

From the Coasian point of view, the answer is twofold. First and 'positively', Coase claims that it does not matter how property rights and liability are allocated as long as they are allocated and provided (unrealistically) that transaction costs are zero.

Coase claims it is wrong to think of the farmer and the railroad as either 'right' or 'wrong' (liable), as 'aggressor' or 'victim':

> The question is commonly thought of as one in which A inflicts harm on B and what has to be decided is, How should we restrain A? But this is wrong. We are dealing with a problem of a reciprocal nature. To avoid the harm to B would be to inflict harm on A. The real question that has to be decided is, Should A be allowed to harm B or should B be allowed to harm A? The problem is to avoid the more serious harm.[12]

Further, given the 'equal' moral standing of A and B, for the allocation of economic resources it allegedly does not matter to whom property rights are initially assigned. Suppose the crop loss to the farmer, A, is $1000, and the cost of a spark apprehension device (SAD) to the railroad, B, is $750. If B is found liable for the crop damage, B will install a SAD or cease operations. If B is found not liable, then A will pay a sum between $750 and $1000 for B to install a SAD. Both possibilities result in the installation. Now assume the numbers are reversed: the crop loss is $750, and the cost of a SAD is $1000. If B is found liable, he will pay A $750, but he will not install a SAD. And if B is found not liable, A is unable to pay B enough to install a SAD. Again, both scenarios end with the same result: there will be *no* SAD. Therefore, regardless of how property rights are initially assigned, according to Coase, Demsetz and Posner the allocation of production factors will be the same.

Second and 'normatively' – and for the only realistic case of positive transaction costs – Coase, Demsetz and Posner demand that courts assign property rights to contesting parties in such a way that 'wealth' or the 'value of production' is maximized. For the case just considered this means that if the cost of the SAD is less than the crop loss, then the court should side with the farmer and hold the railroad liable. Otherwise, if the cost of the SAD is higher than the loss in crops, then the court should side with the railroad and hold the farmer liable. Posner offers another example. A factory emits smoke and thereby lowers residential property values. If property values are lowered by $3 million and the plant relocation cost is $2 million, the plant should be held liable and forced to relocate. Yet if the numbers are reversed – property values fall by $2 million and relocation costs are $3 million – the factory may stay and continue to emit smoke.

Both the positive and the normative claim of Chicago law and economics must be rejected.[13] As for the claim that it does not matter to whom property rights are initially assigned, three responses are in order. First, as Coase cannot help but admit, it certainly matters to the farmer and the railroad to whom which rights are assigned. It matters not just how resources are allocated but also who owns them.

Second and more importantly, for the value of social production it matters fundamentally how property rights are assigned. The resources allocated to productive ventures are not simply given. They themselves are the outcome of previous acts of original appropriation and production, and how much original appropriation and production there is depends on the incentive for appropriators and producers. If appropriators and producers are the absolute owners of what they have appropriated or produced, that is, if no liability *vis-à-vis* second- or third-comers arises out of acts of appropriation and production, then the level of wealth will be maximized. On the other hand, if original appropriators and producers can be found liable *vis-à-vis* latecomers, as is implied in Coase's 'reciprocity of harm' doctrine, then the value of production will be lower than otherwise. That is, the 'it doesn't matter' doctrine is counterproductive to the stated goal of wealth maximization.

Third, Coase's claim that the use of resources will be unaffected by the initial allocation of property rights is not generally true. Indeed, it is easy to produce counterexamples. Suppose the farmer does not lose $1000 in crops because of the railroad's sparks, but he loses a flower garden worth $1000 to him but worthless to anyone else. If the court assigns liability to the railroad, the $750 SAD will be installed. If the court does not assign liability to the railroad, the SAD will not be installed because the farmer simply does not possess the funds to bribe the railroad to install a SAD. The allocation of resources is different depending on the initial assignment of property rights.

Similarly, contra the normative claim of Chicago law and economics that courts should assign property rights so as to maximize social wealth, three responses are in order. First, any interpersonal comparison of utility is scientifically impossible, yet courts must engage in such comparisons willy-nilly whenever they engage in cost–benefit analyses. Such cost–benefit analyses are as arbitrary as the assumptions on which they rest. For example, they assume that psychic costs can be ignored and that the marginal utility of money is constant and the same for everyone.

Second, as the numerical examples given above show, courts assign property rights differently depending on changing market data. If the SAD is less expensive than the crop damage, the farmer is found in the right, while if the SAD is more expensive than the damage, the railroad is found in the right. That is, different circumstances will lead to a redistribution of

property titles. No one can ever be sure of his property.[14] Legal uncertainty is made permanent. This seems neither just nor economical; moreover, who in his right mind would ever turn to a court that announced that it may reallocate existing property titles in the course of time depending on changing market conditions?

Finally, an ethic must not only have permanency and stability with changing circumstances; it must also allow one to make a decision about 'just or unjust' *prior* to one's actions, and it must concern something under an actor's control. Such is the case for the classic private property ethic with its first-use–first-own principle. According to this ethic, to act justly means that a person employs only justly acquired means – means originally appropriated, produced, or contractually acquired from a previous owner – and that he employs them so that no physical damage to others' property results. Every person can determine ex ante whether or not this condition is met, and he has control over whether or not his actions physically damage the property of others. In distinct contrast, the wealth maximization ethic fails in both regards. No one can determine *ex ante* whether or not his actions will lead to social wealth maximization. If this can be determined at all, it can only be determined *ex post*. Nor does anyone have control over whether or not his actions maximize social wealth. Whether or not they do depends on others' actions and evaluations. Again, who in his right mind would subject himself to the judgment of a court that did not let him know in advance how to act justly and how to avoid acting unjustly but that would judge *ex post*, after the facts?

Notes

1. This issue will be developed towards the end of this chapter.
2. See Rothbard ([1962] 1993, [1970] 1977, [1982] 1998, [1974] 2000, 1997).
3. See also Hoppe (1989, 1993).
4. Note the 'natural law' character of the proposed solution to the problem of social order – that private property and its acquisition through acts of original appropriation are not mere conventions but necessary institutions (in accordance with man's nature as a rational animal). A convention serves a purpose, and an alternative to a convention exists. For instance, the Latin alphabet serves the purpose of written communication. It has an alternative, the Cyrillic alphabet. Hence, we call it a convention. What is the purpose of norms? The avoidance of conflict regarding the use of scarce physical things. Conflict-generating norms contradict the very purpose of norms. Yet with regard to the purpose of conflict avoidance, no alternative to private property and original appropriation exists. In the absence of prestabilized harmony among actors, conflict can only be prevented if all goods are always in the private ownership of specific individuals and it is always clear who owns what and who does not. Also, conflicts can only be avoided from the very beginning of mankind if private property is acquired by acts of original appropriation (instead of by mere declarations or words of latecomers).
5. While no one could act if everyone owned the value of his property, it is practically possible that one person or group, A, owns the value of his property and can determine what another person or group, B, may or may not do with the things under their control. This, however, means that B 'owns' neither the value nor the physical integrity of the

things under his control; that is, B and his property are actually owned by A. This rule can be implemented, but it does not qualify as a human ethic. Instead, it is a two-class system of exploiting *Uebermensch* and exploited *Untermensch*.

6. For details, see Rothbard (1995) and also Bethell (1998).
7. Aristotle ([1946], p. 1261b).
8. Ibid., p. 1263a.
9. Ibid., p. 1263b.
10. See Coase (1988); Demsetz (1988); Posner (1981).
11. Posner (1981, p. 74): 'an act of injustice is (defined as) an act that reduces the wealth of society'.
12. Coase ([1960] 1988, p. 96). The moral perversity of this claim is best illustrated by applying it to the case of A raping B. According to Coase, A is not supposed to be restrained. Rather, 'we are dealing with a problem of a reciprocal nature'. In preventing A from raping B, harm is inflicted on A because he can no longer rape freely. The real question is: should A be allowed to rape B, or should B be allowed to prohibit A from raping him/her? 'The problem is to avoid the more serious harm'.
13. See also Block (1977, 1995, 2000); North (1992, 2003).
14. Posner (1981, pp. 70–71) admits this with captivating frankness: 'Absolute rights play an important role in the economic theory of the law. … But when transaction costs are prohibitive, the recognition of absolute rights is inefficient. … property rights, although absolute, [are] contingent on transaction costs and subservient or instrumental to the goal of wealth maximization'.

References

Aristotle [1946], *Politics*, Oxford: Clarendon Press.
Bethell, Tom (1998), *The Noblest Triumph. Property and Prosperity Through the Ages*, New York: St. Martin's Press.
Block, W. (1977), 'Coase and Demsetz on private property rights', *Journal of Libertarian Studies*, **1** (2), 111–15.
Block, W. (1995), 'Ethics, efficiency, Coasian property rights, and psychic income: a reply to Harold Demsetz', *Review of Austrian Economics*, **8** (2), 61–125.
Block, Walter (2000), 'Private property rights, erroneous interpretations, morality and economics', *Quarterly Journal of Austrian Economics*, **3** (1), 63–78.
Coase, Ronald (1960), 'The problem of social cost', reprinted in R. Coase (1988), *The Firm, The Market, and the Law*, Chicago, University of Chicago Press, pp. 95–156.
Coase, Ronald (1988), *The Firm, The Market, and the Law*, Chicago, University of Chicago Press.
Demsetz, Harold (1988), *Ownership, Control, and the Firm*, Oxford: Basil Blackwell.
Hoppe, Hans-Hermann (1989), *A Theory of Socialism and Capitalism*, Boston, MA: Kluwer Academic Publishers.
Hoppe, Hans-Hermann (1993), *The Economics and Ethics of Private Property*, Boston, MA: Kluwer Academic.
North, Gary (1992), *The Coase Theorem: A Study in Epistemology*, Tyler, TX: Institute for Christian Economics.
North, G. (2003), 'Undermining property rights: Coase and Becker', *Journal of Libertarian Studies*, **16** (4), 75–100.
Posner, Richard (1981), *The Economics of Justice*, Cambridge, MA: Harvard University Press.
Rothbard, Murray N. (1962), *Man, Economy, and State*, Princeton, NJ: D. Van Nostrand, reprinted (1993) in Auburn, AL: Mises Institute.
Rothbard, Murray N. (1970), *Power and Market*, Menlo Park, CA: Institute for Human Studies reprinted (1977) in Kansas City, KS: Sheed Andrews & McMeel.
Rothbard, Murray N. (1974), *Egalitarianism as a Revolt against Nature and other Essays*, Washington, DC: Libertarian Review Press reprinted (2000) in Auburn, AL: Mises Institute.
Rothbard, Murray N. (1982), *The Ethics of Liberty*, Atlantic Highlands, NJ: Humanities Press reprinted (1998) in New York: New York University Press.

Rothbard, Murray N. (1995), *Economic Thought Before Adam Smith. An Austrian Perspective on the History of Economic Thought*, vol. I, Aldershot, UK and Brookfield, USA: Edward Elgar.
Rothbard, Murray N. (1997), *The Logic of Action*, 2 vols, Cheltenham, UK and Lyme, USA: Edward Elgar.

3 The origins and evolution of property rights systems

*Francesco Parisi**

Introduction

The institution of property is nearly as old as recorded history. In spite of its stability as a fundamental institution of human society, the concept of property and the privileges, obligations and restrictions that govern ownership have undergone substantial changes throughout history. In this chapter, I shall consider the main stages in the emergence and consolidation of property and discuss some economic theories on the evolution of legal and social conceptions of property.

The history of property is illuminated by economic theory. The concept of scarcity – which some notable authorities believe is at least a necessary condition for the establishment of enforceable property rights – is valuable in explaining the limited domain of property in early societies and the changing contours of property protection as a result of changes in the economic structure of society.[1] As pointed out by Demsetz (1967), property rights develop to internalize externalities in the use of scarce resources. However, there are costs associated with the establishment of property. While scarcity may be necessary for giving objects value and prompting the desire to have property rights, the establishment of such rights also requires that the protection of the rights be economically efficient from a societal standpoint. That is, the marginal benefit of protection (internalization) exceeds the marginal cost of protection. Property rights emerge only when the gains of internalization become larger than the cost of internalization. The study of the historical evolution of property confirms these economic propositions and reveals that changes in the economy often trigger changes in the social and legal conception of property.[2]

Property rights emerge and grow in societies in relation to the cost–benefit calculus regarding the establishment and protection of such rights. Economic change creates new cost–benefit relationships, giving rise to modifications in property regimes (Posner 1998; Rose 1998).[3] This gives rise to a grounds-up conception of property, in which the legal notion of property reflects the localized and evolving function performed by property in society. The natural propensity of humans to possess productive and scarce resources and the social acceptance of this human attitude gives origin to the institution of

property and its regulation. This functional and dynamic evolution of property is the object of this study.

Origins of property

Property, in its ordinary meaning, refers to anything that may be possessed or become the subject of ownership. All societies recognize private property, to a greater or lesser extent. The content of property and its entitlements, however, has undergone substantial change over time. The evolution of the legal and social conceptions of property reveals a close relationship between changes in an economic system and shifts in the structure and content of property rights.

Maine ([1861] 2000) examined whether the property regimes that preceded the modern individual property system evidenced a predominantly communal or individual character. The results of his investigation led him to criticize other theorists' emphasis on the role of individual property. Even when human societies came to accept the idea that the majority of objects can be subjected to private ownership, the holders of such property rights were often family and groups, rather than single individuals. Maine likens the early regimes of property to systems of joint ownership, not separate ownership.[4]

Along similar lines – and in spite of the reluctance of modern social scientists to formulate general theories of evolution – a large number of anthropologists and legal historians have come to agree on the identification of some general patterns in the evolution of property.

The age of hunters

The age of hunters is perhaps one of the first stages in which humans appeared to assert property claims over physical resources. In a hunters' society, property consists mostly of what hunters can kill for their own consumption or trade. At this stage, there is no need to define property rights over other resources, such as land or stock of wild animals, and consequently little need for institutional protection of property rights.

During this era, the social structure is characterized by the presence of tribes, or clans. The modes of production in this phase are hunting and fishing, and joint production and equitable sharing govern the distribution of the bounty.[5] The clan, which is a group of several families, acts, in effect, as a family-based firm.[6] This segment of the tribal property stage appears to correspond closely to Adam Smith's ([1776] 1986) description of the first stage of development in human society (the age of hunters).[7] Smith observes that, at this stage, abundance of unclaimed resources made establishment of property rights unnecessary. Any property claim beyond what the hunters' clan could use for their own consumption would in fact impose unwarranted monitoring and enforcement costs.[8]

Interestingly, Bouckaert (1999) points out that, even when property rights to a stock cannot be established, property rights to its flow might still be created.[9] Such a regime of property claims over flow, rather than stock, of resources, seems to characterize this first stage in the evolution of property.

The age of pasture and shepherds

Economic historians generally identify the second stage in the evolution of human economy as the age of pasture and shepherds. During this period, people begin asserting property claims over animal herds and grazing lands. However, property claims were still asserted by the tribe, rather than by individuals within the group. In most early societies, population increase and gradual competition for the use of land for grazing required that land be divided among tribes: 'The earliest distribution of the land was into pasture and territories of chase common to the tribe, for the idea of individual ownership of the land is of ulterior and tardier growth' (LaFargue 1975, p. 36). Such claims are initially asserted by means of occupation, use or accession. In this phase, tribal property is non-transferable.

As people appropriated herds and flocks, the supply of available animals became scarce, such that the remaining individuals were unable to gain their subsistence from hunting. The property owners consequently grew fearful that the poor would attempt to appropriate their property. This is the stage during which greatly unequal distribution of property first arises. The extremely unequal distribution of claimed resources and subsequent scarcity of unclaimed resources are each necessary, and the combination of them sufficient, for establishment of property rights. Smith ([1776] 1986) argues that individuals thus gained the right to exclude others from their property.

The age of agriculture: from functional to spatial conceptions of property

In subsequent times, agriculture and management of farm animals gradually became the predominant modes of production. Pipes (1991) focuses on resource depletion as the primary genesis of the human notion of property, noting above all else that communal ownership is inefficient. He writes:

> [I]n all primitive societies and most non-Western societies in general, land was not treated as a commodity and hence was not truly property, which, by definition, entails the right of disposal … The transformation of land into tribal, family, or individual ownership seems to occur, first and foremost, in consequence of population pressures which call for a more rational method of exploitation, and it does so because the unregulated exploitation of natural resources leads to their depletion. (Pipes 1991, p. 89)

The transition from hunting and gathering to agriculture around 10 000 BCE made land use more efficient and increased the value and resulting social appreciation of property.[10]

This change in the economy is accompanied by a gradual gain in the importance of the family. Family units gradually acquired interests that were different from those of the clan at large. In this context, the tribal land of the pastoral communities was gradually partitioned among the families that constituted the clan or tribe and what was the communal territory of the tribe was gradually parceled out to become the collective property of individual families.[11] However, the property rights of the family are still subject to regulation by the tribe.

The assignment of land to the family units did not take place in what modern scholars would consider full property. Instead, limited property rights were assigned to family units in order to allow them to carry out more effectively the specific agricultural activity that they intended to perform. The land assigned to family units remained subject to other property claims held by the tribe at large. These included rights for hunting and also compatible grazing uses of the land. The territorial scope of the partial property rights depended on the nature of the rights involved. Thus, for example, the pasture of lands was originally the joint property of all the members of the clan. This was so because given the structure of the economy no single individual or family could have optimally exploited such right to pasture. As LaFargue (1975) points out, the unit of the economy gradually changes from the tribe or clan to the smaller family unit. Gradually the parcels of land were cultivated by each family under the direction of its chief and the supervision of the village council. The resulting crops were the property of the family, and would not become the property of the tribe or clan collectively, as was the case in earlier periods.

In these societies, relatively simple rules governed land ownership. The character of property rights allocated to the family unit was related to the prospective functional use of the land. Other tribe members outside the family could continue to use the land for non-agricultural purposes. For example, those who used the land to hunt continued to hold hunting privileges, and those who raised livestock could hold grazing rights in the same geographic area. Such functional conceptions of property were the natural consequence of the derivation of property from actual use and possession of the land. Over time, the way in which the land was used determined the kinds of possession possible.[12] This system often resulted in multiple property claims coexisting on the same land. Customary rules then regulated the possession, use and transfer of such functional rights. Such functional divisions often made good sense because different owners could undertake specialized activities over the same territory with little encroachment on one another.

Given the low population density and the limited rate of exploitation of natural resources, functional partitions of property were indeed often efficient. They provided an opportunity to allocate the same land towards multiple privately held use rights, allowing an optimal level of exploitation, as all parts of the property could be used.[13] Detailed customs, based on past usage and historic rights, determined what was considered acceptable conduct with respect to the interaction (natural externalities) between the various activities. Early societies thus embraced a 'functional' conception of property, in which property rights were related to specific uses of the land, rather than a spatial conception of property in which the confines of property were determined by physical boundaries. Property was not divided along spatial lines, as in the modern world, but through horizontal functional partitions, in which different individuals or families would own specific rights over the land related to specific uses (for example, farming, fishing, hunting and so on).

As time progressed, however, agricultural societies developed a more complex conception of property in which functional partitions of rights survived as exceptions to a regime of unified ownership.[14] This paradigmatic shift is understandable, given that in an agricultural economy the coexistence of multiple rights over the same land created conflicts and increased opportunities for wasteful externalities. Furthermore, functional partitioning of land, while efficient in a stable economy, became unsustainable in conditions of rapid economic change. Moving from pastoral to agricultural economies, many societies thus changed their property systems, abandoning functional property in favor of spatial property (that is, making their property systems more similar to those we are accustomed to observing in the modern Western world).

This transition has a plausible economic explanation. With a rapidly changing economy, optimal uses of land are also subject to rapid flux. A functional property regime impeded the transformation of land to optimal use, given the wide range of rights that had to be accommodated or superceded.

In spatial property regimes, a single owner generally holds all rights pertaining to a tract of land. Such unified ownership could better serve the needs of a changing economy. The division of property along functional lines, while allowing the optimization of property with respect to all of its potential uses, did not provide sufficient flexibility to accommodate structural transformations over time.

The Roman law concept of property is thus based on the realization that the optimal use of the land is subject to change over time, and that absolute ownership rights provide greater flexibility, given the concentration of decision rights in the hands of a single individual.[15]

The fall and rise of absolute property

The historical evolution of property led to the emergence (and gradual dominance) of spatial conceptions of property, with absolute property regimes as the default legal rule in all established agricultural societies. During the feudal era, a new array of functional and legal limitations on the use and disposition of land encroached on the Roman conception of absolute property. Although the foundations of the medieval law of property were unquestionably Roman, the feudal system gradually transformed the accepted social conception of property.

Physical unity and legal disunity in medieval land law

In the feudal world, rights and duties were based on land tenure and personal relationships, and this conception of property was instrumental in maintaining feudal social and economic structure. The early types of land licenses resembled grants of full ownership, but in later times the kings and the lesser lords kept the ownership of the land to themselves and granted only partial rights of use and exploitation.[16] Land was held in fief by vassals as a result of a grant by their lords in exchange for a variety of services and vows of personal loyalty.[17] These grants of fragmented ownership gradually became hereditary holdings.[18] Customary norms prevented the unilateral abrogation of these grants, except as a result of legal forfeiture and seizure.[19] This resulted in a multilayered, and potentially irreversible, fragmentation of property.

In this feudal system of land tenure, each individual was defined by his hierarchical status and relationship to land. With the sole exception of the king, every individual was subservient to another. According to the well-known feudal pyramid, only the lesser tenants had possessory use of the land, and all the others served as intermediaries in the collection of fees and granting of services and protection. The king stood as the ultimate residual claimant. Through this process, feudal property became quite distinct from the Roman paradigm of property, as feudal grants were always limited by the act of license and title; possessory interests never resided in the same hands. Property ownership was neither unlimited nor absolute; interests were not enforceable *erga omnes* (toward all), but rather consisted of a bundle of rights and duties, partially applicable to the whole community and partially determined by the specific contractual relationship between the grantor and the grantee. A complex system of political and social control reinforced this transition from the Roman system to the feudal regime of dispersed ownership (and property fragmentation).

Feudalism was inextricably linked to agricultural life.[20] In an agricultural economy, functional forms of fragmentation are generally not problematic, as long as the physical unity of land is preserved. In this respect, feudal legal systems were designed to limit the risk of excessive fragmentation.[21] Rules of

primogeniture[22] and prohibition of subinfeuds[23] are examples of the attempts of feudal law to constrain entropy in property.

The peculiar coexistence of physical unity and legal disunity in feudal property could have functioned properly in an agricultural economy, but proved problematic in any other kind of economic context. Indeed, feudal arrangements, which generally flourished in closed agricultural economies, did not take root in urban environments.[24] The cities of the Roman Empire, in as far as they survived at all, did not have a parallel feudal structure,[25] with the exception of some urban areas in Italy and southern France.[26]

The rise of absolute conceptions of property

With the approach of the modern era came another paradigmatic shift in the conception of property. Just as the transition from pastoral to agricultural economies rendered the so-called functional conceptions of property impracticable, the gradual growth of the economy made the feudal dispersion of control over property highly problematic.

As generally recognized in the literature, the abolition of feudalism was a necessary precondition for the shift to a modern market in land, in which individuals can transfer full ownership and development rights to third parties through contracts or testamentary dispositions. The transition back from the relativistic and contractual basis of property to the Roman absolute conception of ownership, was not gradual and smooth, however.[27] The historical events surrounding the end of the feudal era demonstrate the power of the irreversible dynamics of entropy in property, a theme of this chapter.

Both a political and an ideological revolution were required to reshape the dominant conception and content of property. The French Revolution marked a critical turning point. In a vain attempt to stave off serious political and social upheaval, the French nobles and clergymen renounced some of their feudal privileges at the first session of the *États généraux* (States-General) in over 200 years on 4 August 1789. The theoretical significance of this action was great, as it freed the land from a multitude of personal servitudes (for example, hunting rights and labor services). Other feudal burdens (for example, seigneurial fees) could be extinguished by paying a lump-sum amount to the lords (generally corresponding to 20 to 25 times the value of the annuity). In practice, the peasantry was unable to pay the large amounts required to obtain release from the feudal servitudes. The French Revolution itself brought with it the collapse of the feudal regime, with the outright abolition of all burdens and seigneurial fees without compensation.

During the eighteenth century, it had become fashionable to point to the feudal tradition as the root of the inefficient property fragmentation and to rebel against the feudal heritage by proclaiming a new paradigm of absolute and unified property. Historically, feudal land systems imposed on landholders

positive obligations, which were not part of the Romanistic bundle of rights and duties of property holders, mixing absolute and relative rights in a hybrid property relationship. Hegel's ([1821] 1942) philosophical version of this history has influenced the legal conceptions of property around the world. Hegel purports that the standardization of property interests is a movement that is related to the difficult struggle to free property from the pervasive feudal encumbrances, suggesting that individual freedom closely depends on the freedom of property.[28]

In this setting, the Roman approach became the model for bourgeois property, conceived as an absolute private right to the enjoyment of one's land. This renewed conception is at the origin of much theoretical work in legal theory and philosophy. One can think of John Austin's ([1832] 1885, p. 808) premise that the right of property consists of two main elements: the right to use the property and the power to exclude others. Most importantly, eighteenth- and nineteenth-century theorists specified the structural attributes of a property right. Among others, we find the Kantian notion that universal norms must be negative in content, that is, they 'must command individuals *not* to do something' and that 'the correlative of a real right cannot require action'.[29]

This conception of absolute property contrasted dramatically with the concept of feudal property as a bundle of rights and duties that mixed relative and absolute relations in both private and public spheres. Although, in many respects, such a unified conception of property was already present in the pre-feudal world, the intellectual reaction against the old regime led to a more nuanced articulation of this ideal.[30] This theoretical evolution culminated in the revolutionary events of the 1790s that marked the beginning of a new era in the property regimes of France and the rest of Europe. This new legal approach, in order to meld successfully the concepts of functional unity with the revived absolute conception of property, had to have substantive rules to foster functional, physical and legal unity in property. These principles of unitary property are embodied in several important rules that characterize modern property law, which I shall consider next.

Structural variations in property and the foundations of absolute property

As a reaction to the feudal tradition, the rationalist jurisprudence of the eighteenth century and the modern codifications of the nineteenth century revived the various important Roman rules of property, recasting them as general principles of civil law. The principles of unity in property can be tentatively grouped under the headings of (i) functional, (ii) physical, and (iii) legal unity. These related principles contribute in different ways to control the problems of entropy in property.

Principles of functional unity in property

Under classical Roman law, the property owner (*proprietarius*) was not allowed to transfer anything less than the entire bundle of rights, privileges and powers that he had in the property. Conveyances of rights in a lesser measure than full ownership were only permitted on an exceptional basis and in a limited number of cases.[31] Thus, for example, the creation of legally binding restrictions on property was limited to situations in which the dominant estate could demonstrate a perpetual need for the arrangement. In the Roman Digest we read that servitudes necessitate a *causa perpetua*.[32] In other passages, the Roman sources explicitly indicate that the servitudes created for the transitional benefit of the owner of a neighboring lot (as opposed to the perpetual benefit of the land itself) were not valid.[33]

As I have discussed above, the notion of absolute ownership underwent a substantial change in feudal law, but eventually regained popularity at the time of the modern European codifications. The modern codes limit the permissible level of functional property fragmentation and further provide property-type protection only for specific, socially desirable, property rights.[34] This favoring of certain property arrangements is known as the *numerus clausus* principle, and is an important expression of the fundamental principle of unity that underlies modern property law. The purpose of this principle is to forestall private individuals from creating property rights that differ from those that are expressly recognized by the legal system.[35]

The early formulations of the *numerus clausus* lacked a well-articulated rationale, especially striking because it contrasts sharply with the doctrine of freedom of contract, namely that two parties to a private contract may agree on virtually any arrangement without government limitations.[36] The dichotomy between these contract and property paradigms results in a general tension between the principle of freedom to contract and the societal need for standardization in property law. The modern European codifications all reflect this tension. They promote freedom of contract by recognizing and fully enforcing, both nominate and innominate forms of contract. Yet at the same time they limit private autonomy in property transactions and only enforce transactions pertaining to standardized (or nominate) forms of property.[37]

Although in many ways the intellectual product of the French Revolution, the influence of the *numerus clausus* principle has lasted well beyond the post-revolutionary codes, and can be found in most of the modern European codes. The French *Code Napoléon* (Napoleonic Code) of 1804,[38] the German civil code *Burgerliches Gesetzbuch* (BGB) of 1900,[39] and several other codifications[40] contain provisions that restrict the creation (or at least withhold the enforcement) of atypical property rights. As Rudden (1987, p. 243) aptly put it, ‘in very general terms, all systems limit, or at least greatly restrict, the creation of real rights: “fancies” are for contract, not property’.

These limits on the creation of atypical property rights eventually emerged as a general principle of modern property law. Even jurisdictions that have not formally codified the doctrine adhere to its strictures.

The requirement that the transfer of immovable property be recorded in a public registry enforces the *numerus clausus* principle because only nominate property rights can be duly registered, thereby ensuring the intelligibility of public records for notice and publicity purposes.[41] It follows logically that any contract which constitutes or modifies a property situation in contempt of the taxonomy of real rights recognized by the legal system is only a source of contractual obligations.[42] But Gambaro (1995, p. 67) illustrates the paradox of allowing parties to enter into binding contractual obligations, yet preventing them from enjoying the benefits of those agreements. If the public record system does not allow atypical property rights to be recorded, freedom of contract is itself undermined, because the system withholds the mechanism, namely the recording system, that could transform an atypical property agreement between the parties from a personal obligation into a real right enforceable against third-party purchasers.

Principles of physical unity in property

Other basic principles of modern property law demonstrate the general tendency of legal systems to combat entropy and promote unity in property. In dealing with the physical partition of property, the rules of civil law systems symbolize the ideal of physical integrity of property. For instance, a large number of civil law systems jurisprudentially or legislatively recognize the owner's right to fence property as a symbolic prerogative of his/her sovereignty.[43] Furthermore, civil law systems address the problem of physical unity with rules restricting horizontal partitions of property building on the heritage of Roman legal systems that generally limited recognition of subsoil real rights, as suggested by the Latin maxim: '*Cuius est solum eius est usque ad coelum et usque ad inferos*' (whoever owns the land owns the property all the way to heaven and all the way to the center of the earth).[44]

The theory and practice of property law regarding physical unity has undergone several changes over the centuries. Whenever recognized, the *ad coelum* rule presumes that someone who buys property unaware of any obstacle to its free use (bona fide purchaser) acquires priority over other claims on the land, be they for underground or surface resources.

In spite of the medieval departures from unified conceptions of property, most of the modern civil codes of the nineteenth century reinstated the Romanistic conception of property disallowing horizontal forms of property fragmentation. In both the French Code of 1804 and the Italian Code of 1865 land could not be horizontally severed into multiple surface and subsurface estates and legal title to the various land strata had to vest in a single owner.

Early common law erected similar obstacles to the horizontal fragmentation of property[45] and, thanks to the work of Lord Coke in the early seventeenth century, the abbreviation '*ad coelum*' became an established legal concept in English law.[46]

These constraints on the freedom of the parties were an important corollary of the principle of physical unity in property that led to the eighteenth-century intellectual reaction against feudal property fragmentation, but they did not shape the ultimate approach to property that emerged in Continental Europe. In spite of the *numerus clausus* principle and the other formulations of the ideal of unitary property, the explicit prohibitions of the modern codes proved ineffective.[47] Property owners continued to partition their land into multiple surface and subsurface estates. Originally such agreements could not convey real title to the various land strata, but parties occasionally attempted to bypass this impediment by agreeing not to invoke accession rules against the titular owner of surface rights, should the informal surface owner decide to erect a building on the land.[48] The courts were initially reluctant to enforce the parties' agreements, which they found in open contravention of the unity rule. Over time, however, civil courts developed a more accommodating attitude and allowed such atypical forms of property fragmentation to survive in the shadow of the law.

The twentieth-century codes eventually abandoned the rule prohibiting the horizontal fragmentation of property. Starting with the German BGB of 1900[49] and the Italian Civil Code of 1942,[50] civil law moved away from applying the principle of physical unity, reverting to the standards in effect prior to the modern codifications.

Any one or more of three practical reasons discussed below reversed the trend. First, horizontal fragmentation was so commonly tolerated that it was no longer exceptional, and unity became a symbolic legal fiction. Enforcing a rule of unity under such circumstances risked disrupting a peaceful status quo in order to confront the unavoidable dilemma of deciding which of the two good faith parties should acquire title to the property.[51] Second, the risks of horizontal forms of property fragmentation are limited: few parties engage in such partitions, and in practice no more than two layers – surface and subsoil – are likely. This limited form of fragmentation does not raise serious strategic problems, given the bilateral monopoly of the two fragmented owners under the circumstances. Third, the practical need for regulating mineral rights and rights in the exploitation of underground resources prompted the gradual abandonment of the older dogma.

Mid-twentieth-century civil law scholars criticized abandoning the rule of physical unity because doing so violated the modern ideal of unified property. Horizontal property fragmentation appeared antithetical to traditional notions of property ownership that created in a landowner absolute, indivisible rights

to a vertical space extending *usque ad coelum*, *usque ad inferos*, as well as creating mutual constraints on surface and subsurface ownership.[52]

Principles of legal unity in property

A third principle of Western law, granting the owner absolute power to dispose of his/her property, is also closely related to the concept of unity in property, although with quite different implications. All the main European codes enunciate this principle. For instance, Article 544 of the French Civil Code states 'Ownership is the right to enjoy and dispose of things in the most absolute manner', a provision included almost verbatim in the Italian Civil Code of 1865. Similarly, Paragraph 903 of the German BGB affirms that the owner of an object 'may deal with the thing as he pleases and exclude others from any interference'.

While these statements appear uncontroversial at first glance, they become difficult to implement with joint ownership. The principle of absolute disposition indeed becomes an oxymoron when two or more individuals jointly hold decision rights. In addressing this problem, legal systems have historically adopted rules that (a) facilitate the reunification of use and (b) place exclusion rights in the hands of a single individual. The common law achieved these two objectives by making it difficult to create joint tenancy (a legal fiction in which two or more people are regarded as a single owner) and relatively easy to destroy the arrangement. In order to create a joint tenancy, the owners had to be able to demonstrate the four 'unities', namely: (a) time (they acquired the property at the same time); (b) title (they all signed the same instrument); (c) interest (they owned identical rights); and (d) possession. If any of these elements was missing, then the joint tenancy could not be created.

Along similar lines, anyone who found him/herself owning something jointly with others could cause the common property to be divided, in keeping with the Latin maxim *nemo invitus ad communionem compellitur* (no one can be forced to have common property with another). Division could be done unilaterally. All one joint tenant had to do was convey his/her interest to a third party, and the joint tenancy was severed, reverting to a tenancy in common and allowing owners to convey or devise their interests to third parties. Some early cases found that merely expressing the intent to sever the joint tenancy was sufficient to do so. If the tenants could not agree on the management of the property, another option was to petition for partition, in which case the court would either divide up the property or order it sold and divide the proceeds among the owners. Very similar rules are present in civil law jurisdictions to minimize the hold-up power of joint owners in the use of the joint property. In application of the principle of legal unity in property, these systems introduced mechanisms that were easily triggered to allow

owners greater autonomy in disposing of their property, even when others had rights over the same land.

Towards the recognition of new forms of property

As discussed above, in the modern era, legal systems around the world have in different ways manifested a general reluctance to recognize atypical property agreements as enforceable real rights.[53] In recent decades, however, courts and legislatures in both civil and common law jurisdictions, attuned to the modern needs of land developers and property owners, have recognized new property arrangements.[54] The clearest example of this gradual expansion of standard property arrangements in civil law jurisdictions can be found in the area of covenants that run with the land and that occasionally create new *sui generis* real rights. A real covenant is a promise to do, or refrain from doing, something that is connected to land in a legally significant way.[55] Under traditional common law, the rights and duties associated with contracts were not assignable (Corbin 1926) because parties to the original agreement did not have the right to bind third parties to adhere to their arrangement. Accordingly, the benefits and burdens of the original covenants did not transfer with the interest in the land. In many situations, this frustrated the purpose of creating a real covenant in the first place.

Due to the perceived net benefits in having the rights and burdens of a real covenant attach to the title to property (that is, 'run' with the land), courts gradually created a new body of law to overcome the obstacles posed by traditional property and contract theories.[56] Almost without exception, however, legal systems implementing these innovations have created atypical regimes to govern remedial protection and regulate these new rights, rules that diverge substantially from the traditional principles governing property or contracts. Commentators generally attribute these divergences to mere historical accidents (Yiannopoulos 1983; Dwyer and Menell 1998, p. 760). Contrary to the common wisdom in the literature, I suggest that these anomalies are not haphazard.

In order to protect these newly recognized real rights, courts have developed an elaborate set of requirements to minimize the long-term effects of the non-conforming fragmentation of property, adopting a set of rules that differ from traditional property or contract law. Legal systems instead balance the need to mitigate entropy in property by creating perpetual restrictions on the use and alienability of property with the demands of landowners and property developers, wishing to exercise their contractual freedom to dispose of their property as they deem appropriate. Various legal traditions have employed different instruments to achieve this goal. For example, under modern French law, courts do not recognize atypical property covenants as sources of real rights, though they allow parties to approximate a real right by

drawing on the notion of transferable obligations. Thus, French cases have construed contracts between property owners as sources of obligations that are effective against third persons.[57] In Germany and Greece, atypical property covenants are also not enforced as real rights, but, as Yiannopoulos (1983) points out, allowing the contractual remedies to extend beyond the original parties to the covenant produces similar effects.

Interestingly, legal systems often encourage open access to common property (for example, roads, navigation, communications, ideas after the expiration of intellectual property rights and so on),[58] and in other cases the legal system creates and facilitates fragmentation. For instance, the social planner uses entropy to his/her benefit by using conservation easements and the fragmentation (for example, multiplication) of administrative agencies overseeing of land development to slow the pace of suburban development.[59] In yet other instances, the owners themselves structure the non-conforming property arrangements.[60]

Although problematic as a rule, non-conforming partitioning of property rights may be somewhat sensible in achieving specific policy goals or other objectives that property owners desire. These idiosyncratic arrangements are both a reflection of the individual's right to freedom of contract and a legitimate policy instrument for the urban planner. In sum, respecting individual autonomy while minimizing the undesirable deadweight losses that could result from these arrangements is the critical goal.

Conclusion

In this chapter, I have traced the origin and evolution of property rights regimes in light of the changes to economic systems since the earliest days of human civilization. The foundations for this study lie to a certain extent in the work of Demsetz (1967) and other philosophical minds such as Smith ([1776] 1986) and Hegel ([1821] 1942), who recognize that when resources are scarce, human societies formulate property rights to allocate use and regulate production. The development of property rights over time is nothing if not dynamic. The early stages of property – the ages of hunters, pastures and shepherds, and agriculture – reveal the origins of some central tenants of modern property such as commonality (that is, sharing mechanisms), the right to exclude, customary restrictions and spatial property notions. In the feudal era, the juxtaposition of physical unity and legal disunity rapidly propel Western societies towards revolutionary changes in property rights. Absolute rights emerge from the feudal era as individuals gain rights to use and transfer land. However, in the premodern era, Western societies struggle with the extent to which functional, physical and legal unity ought to restrict an individual's bundle of rights in property.

Finally, in modern times, economics remains useful to examine the continuing changes in property rights regimes. Covenants, used frequently by land developers, have emerged as a medium for owners to exercise contractual freedom yet preserve unity. As the modern economy changes, so too will Western systems of property.

Notes

* I would like to thank Lee Istrail and Peter Irvine for their valuable research and editorial assistance.

1. See Smith ([1776] 1986); Demsetz (1967). Rose (1985) similarly suggests that in addition to scarcity, 'we need the capacity to shut out others from the resources that are the objects of our desire, at least when those objects become scarce' and that 'by allocating exclusive control of resources to individuals, a property regime winds up by satisfying even more desires, because it mediates conflicts between individuals and encourages everyone to work and trade instead of fighting, thus making possible an even greater satisfaction of desires' (Ellickson et al., 1995, p. 22).
2. Demsetz (1988, p. 107) pointed out that 'Increased internalization, in the main, results from changes in economic values, changes which stem from the development of new technology and the opening of new markets, changes to which old property rights are poorly attuned'.
3. Posner (1998, p. 36) points out that, under common law, domestic animals are owned like any other personal property, whereas wild animals are not owned until caught or killed. This is because the cost of enforcing private property rights over wild animals outweighs the value of the animals.
4. Looking at the origins of property in Chapter 8 of his book, Maine ([1861] 2000) thinks a major flaw in the occupancy theories of the origin of property is that they look at individuals rather than families and groups. He says it is likely that property originated as a communal claim. He looks at the village communities in India to support this theory. In these communities, as soon as a son is born, he acquires a vested interest in his father's property (the family estate). The property remains undivided for several generations.
5. LaFargue (1975, pp. 22–4) uses data from Lewis Henry Morgan's (d. 1881) famous anthropological studies to suggest that a similar phase of evolution characterized the social setting of Native American tribes. These clans shared in the product of each producer, distributed meals equitably, ate moderately, and the tribe's life could be characterized as 'primitive communism'.
6. LaFargue (1975) defines a family as a 'consanguine collective', or a group of blood relations.
7. Adam Smith ([1776] 1986, pp. 69–70) identifies four stages in the development of civil society, two of which are closely tied to the origins of property. The first stage, according to Adam Smith is the age of hunters.
8. Later economists have rationalized the conditions for the emergence of property rights. See Demsetz (1967 and 1988), showing that scarcity may be a necessary but not a sufficient condition for the establishment of private property rights.
9. Bouckaert (1999) discusses various property rules that maximize societal utility, drawing a distinction between a property right to a stock and a right to its flow. For example, assigning property rights to own a herd of wild animals would be prohibitively expensive but rights to captured and killed game can be created efficiently.
10. 'Hunting and gathering, though involving relative little effort, is exceedingly wasteful of land' (Pipes 1991, pp. 92–3).
11. In addition to parents and their children, the family unit also included the father's 'concubines … his children, his younger brothers, with their wives and children, and his unmarried sisters' (LaFargue 1975, p. 50). 'The arable lands, hitherto cultivated in common by the entire clan, are divided into parcels of different categories, according to the quality of the soil … the number of lots corresponds to that of the families' (ibid., p. 52).

12. No single act of possession can encompass all potential forms of land use, meaning that the system of deriving ownership from possession generates limited (or 'functional') property rights (Rose 1985; Parisi 2002a).
13. For an application of this framework to modern settings, see Banner (1999) and Henry Smith (2000).
14. Absolute property rights can be observed only as an exceptional category in most non-Western societies while they represent a default property regime in modern Western legal systems. Historically, the absolute Western conception of property is not universal.
15. Obviously, if a single owner can claim a tract of land, a different fragmentation problem may take place: excessive spatial fragmentation, leading to inefficiencies of scale. Western agricultural societies dealt with this problem in a variety of ways. In Roman law, the head of the household (*paterfamilias)* had concentrated authority over the property. An elaborate system (*peculium*) mitigated this concentrated legal capacity as with other members of the family group, such as slaves and sons, authorized to make binding legal transactions relating to property. In later times, when legal capacity was extended to every individual of majority age within the group, the fragmentation of property was prevented though other rules and social customs, such as succession rules (for example, rules of primogeniture), and institutional arrangements (for example, feudal hierarchies) resulting in the concentration of land in the hands of a few individuals.
16. Other privileges of the lord included the so-called 'feudal incidents', which, among other things, gave the lord the right to possess the land.
17. Over time, the services, which were originally related to supporting and defending the lord in time of war through military service were gradually converted into pecuniary obligations. It is impossible to understand the developments of medieval society without realizing that the Crown and the nobility (and within each major feudal manor the lord and his vassals) were power centers that were always potentially in conflict. The famous feudal pyramid (king, nobility, lords, vassals) depicts a dynamic society that was often on the verge of disintegration. On the personal level, the loyalty agreement between a lord and his vassal always tended to mask that same kind of unstable relationship and, likewise, the property ties attempted to create a bonding mechanism that would foster stability.
18. If a tenant died without heirs, the land returned to the lord: a form of residual claim of the sovereign that survives, under a different name, in the modern law of succession.
19. Forfeiture and seizure were remedies that allowed the lord to regain possession of the land if a tenant had breached his oath of loyalty or failed to perform the applicable feudal services. Similar remedies applied in the case of high treason in favor of the king. For further discussion, see Dukeminier and Krier (2002).
20. The feudal economy was largely based on agriculture. The feudal structure ensured the production of food needed for maintaining a population constantly on the verge of war.
21. Even the feudal law of property – often presented as the paradigm of entropic property – conceived remedies to combat excessive property fragmentation. Indeed, while some functional forms of fragmentation of property were instrumental to the stability of the feudal society, others could be easily prevented.
22. In the Western legal tradition, laws forbidding the partitioning of land and establishing the succession to land in favor of the youngest or eldest son have often been utilized to preserve the unity of land. These rules often had customary origins and enjoyed a large degree of voluntary compliance, given the interest of most landowners in protecting the power and prestige of the family, which was traditionally linked to the size of land holdings. In this context, priority in succession was traditionally given to the eldest son (primogeniture) or to the youngest son (ultimogeniture). The principal effect of these rules has been the maintenance of unity in the estate of the deceased.
23. Rules emerged to prevent the further fragmentation of possessory interests even in feudal times. The tenants in demesne could not subcontract their rights and obligations through the creation of a lower rank of feudal agents. The practice of subinfeudation has provided much of the intellectual justification for the modern constraints on functional property fragmentation. Rudden (1987) refers to this idea as the 'pyramiding' rationale: if each succeeding owner of a property interest has the power to intertwine the land with the

performance of a fancy, then the land will be encrusted with a pyramid of obligations not unlike those that were associated with the practice of subinfeudation. Rudden acknowledges the weight of this argument, but makes three objections. First, he thinks that the historical objection to subinfeudation did not primarily concern the layers of obligation. Second, he thinks that a legal mechanism could check the excesses of a wider system of property interests. Third, market mechanisms rather than legal mechanisms may better solve such pyramid problems.

24. According to Monateri (1996), it is not coincidental that the rebirth of Roman legal studies began at a time of revival of urban life in northern Italy. Southern France was the center of a code-oriented legal system, as the Roman law remained in force. In contrast, a more complete feudal system in northern France led to the development of customary law under the Carolingian Franks. The cities were not the centers of feudal, but rather ecclesiastical, power. The canon law took precedence, although the merchant law was an important component of the *ius commune*, at least in localities with active commercial markets. Parisi and O'Hara (1998) and Parisi (1998) observe that although jurisdictional conflicts were possible, feudal law harmoniously coexisted with the merchant law, given the different focus on property and contract issues (that is, feudal law generally covered property and status issues, while the latter mostly dealt with contract and commercial issues).
25. The feudal system did not form the base of jurisdiction in cities, where canon and Roman laws took precedence. Feudal law was not congenial to the dynamic needs of the emerging mercantile class. As a result of the commercial revolution brought about by the medieval mercantile class, the large cities of the former Roman Empire became attractive localities for the development of active commercial markets, and most second-generation medieval merchants relocated their center of activity in large cities. At this point, the law merchant became an increasingly important component of the *ius commune.* For further analysis, see Galgano (1976).
26. Feudalism spread from France to northern Italy, Germany and Spain and, later, into some of the eastern Latin territories of Europe. Most of the other great civilizations of the world have gone through periods resembling the feudal arrangements.
27. The feudal property arrangements characterized much of the customary law of property, but medieval academic jurists continued to utilize the Roman categories of property in their scholarly writings. This facilitated the return to the Roman categories that were eventually reinstated as part of the opposition to feudalism that helped lead to the French Revolution.
28. Rudden (1987, p. 250) points out that 'the word "servitude" covers both slavery and easements'.
29. Rudden (1987, p. 249) thinks that these logically relate to Austin's premise, citing Immanuel Kant ([1797] 1887, p. 14) in support of these propositions.
30. Mattei (2000, p. 14) observes that the modern unitary theory of property rights is indeed the intellectual product of the French Revolution. Along similar lines, the extensive work of Rodotà (1990) demonstrates the limits of the analogies between the Roman notions of absolute property and the modern restatement of such notion contained in Article 544 of the French Code of 1804. For further historical analysis of this important transition in the conception of property, see Monateri (1996).
31. Thus, for example, use and exploitation rights divorced from ownership (*usufructus*) could be given only to a living person for the duration of his lifetime; the creation of legally binding restrictions on property (*servitudes*) was sharply limited.
32. Paulus Book 25 *ad Sabinum* in D. 8.2.28: '*omnes autem servitutes praediorum perpetuas causas habere debent*' (all servitudes must have a perpetual cause).
33. Paulus Book 15 *ad Plautium* in D. 8.1.8: '*ut pomum decerpere liceat et ut spatiari et ut cenare in alieno possimus, servitus imponi non potest*' (servitudes cannot be created to grant rights for harvesting fruit or to have meals or merely to walk on another's property). Such atypical arrangements – it was understood as an implicit corollary – could, however, be created as a matter of personal obligations.
34. European scholars also refer to this principle by invoking the concept of nominate property rights. Merrill and Smith (2000, p. 69) have recognized that, although the *numerus*

clausus principle is mostly a Roman law doctrine followed and enforced in most civil law countries, the principle also exists as part of the unarticulated tradition of the common law. The authors illustrate the many ways in which common law judges are accustomed to thinking in terms comparable to the civilian doctrine.

35. For a modern challenge to the *numerus clausus* principle, see Rudden (1987) who critically analyses the legal, philosophical and economic justifications for limiting the types of legally cognized property interests to a handful of standardized forms. See also Parisi (2002b) and Schulz et al. (2002).
36. Merrill and Smith (2000, pp. 68–9) note the peculiar dichotomy between property and contracts, observing that while contract rights are freely customizable, property rights are restricted to a closed list of standardized forms.
37. This implies that property rights are only enforced with real remedies if they conform to one of the 'named' standardized categories. Conversely, the presumption is the opposite in the field of contracts: the legal system enforces all types of contracts unless they violate a mandatory rule of law concerning their object and scope.
38. Several articles of the French Civil Code embrace the concept of 'typicality' of real rights and articulate principles of unitary and absolute property. See, for example, Article 516 on the differentiation of property; Article 526, enlisting the recognized forms of limited real rights (usufruct, servitudes and mortgages); Articles 544–6 on the definition and necessary content of absolute ownership and so on.
39. BGB Paragraph 90, by providing that 'Only corporeal objects are things in the legal sense' can be seen as a substantial departure from the feudal conception of property, where most atypical rights had an intangible nature.
40. Practically all important modern codifications – not all of which were directly influenced by the French and German models – embrace a similar principle of unity in property. Rudden (1987) provides a comparative survey of the *numerus clausus* principle in the modern legal systems of the world, reporting that many Asian legal systems have adopted a basic rule according to which 'no real rights can be created other than those provided for in this Code or other legislation', for example, Korean Civil Code (CC) 185, Thai CC 1298 and Japanese CC 175. Similar provisions exist in other systems of direct European derivation such as Louisiana CC 476–8, Argentine CC 2536, Ethiopian CC 1204 (2), and Israeli Land Law 1969 sections 2–5.
41. The 'absence of notice' legal rationale purports that it would be difficult for a purchaser to know about 'fancies', or non-*numerus clausus* property interests. Rudden (1987) objects to this rationale. He thinks that a functioning recording system could reveal fancies to a purchaser. Further, notice is neither necessary nor sufficient to create valid property interests. For further analysis of this point, see Mattei (2000, pp. 91–2), who observes that the same restriction on the admissibility of recordable instruments does not exist in common law jurisdictions, where parties can create property rights with a much larger degree of autonomy (for example, by means of the creation of a trust instrument).
42. Rudden (1987, p. 243) notes that Argentina is the only country that articulated this important logical corollary in the form of a code provision: Argentine CC 2536.
43. Paragraph 903 of the German BGB, granting the power of absolute disposition, similarly recognized by other codes and civilian courts. See also Mattei (2000, p. 123).
44. The '*ad coelum*' rule was first found in Gaius' *Institutiones*, and later reproduced in Justinian's Digest and *Institutiones*. The basic concept is that real property extends vertically all the way to hell (*inferos*) and heaven (*coelum*). Already in the Republican era of Roman law, however, the so-called grants *ad aedificandum* (which were later given full remedial protection through the *actio de superficie* allowed the creation of horizontal surface rights that could be held by subjects other than the owner of the subsurface estate.
45. Wenzel (1993) points out that the horizontal fragmentation of property was not permitted under early common law. She attributes this to two main reasons: (a) the popularity of the *ad coelum* rule in medieval English thinking and the dominance of absolutist theories of property; and (b) the practices of land transfer that, through the ritual of 'seisin' (resembling the *traditio simbolica* of the Roman law), required the symbolic conveyance of possession via the manual delivery of a stone or clump of soil taken from the land. The

ritual of the *traditio simbolica*, Wenzel suggests, was not suitable to symbolize the transfer of title to subsurface rights or undiscovered mineral rights.

46. The *ad coelum* doctrine is still used in modern law in two main settings, one being in property law, to approach questions of adjoining ownership (for example, cases of constructions hanging over the neighboring land and so on); and in international law, to approach issues of territorial sovereignty over airspace. Most recently, the doctrine has been revived in public international law through the claims of equatorial states for their rights over the geostationary orbit (for satellites). Obviously, non-equatorial states oppose the applicability of the *ad coelum* rule and invoke the application of a rule of first possession.
47. As to the vertical limits of property, even in the early times, the appeal to the *ad coelum* rule was mostly symbolic and subject to several exceptions. The value of such symbolism, however, should not be underestimated, given the interpretive force that is often associated with general principles in civil law systems.
48. According to the accession rule, *superficies solo cedit*, absent such agreement, the owner of title could claim ownership to any construction erected on the land by third parties.
49. BGB, paras 1012–17, which were later replaced by a special law of 15 January 1919, which provided a more explicit regulation of the matter.
50. Article 952 of the Italian Civil Code of 1942 recognizes surface rights as an enforceable real right.
51. Conflicting claims of building- and landowners would have been resolved according to the civilian rules of accession, according to which the owner of the land would acquire title to the building erected by third parties (that is, *superficies solo cedit*) with a duty to compensate the latter for the lesser amount between the incremental value of the property and the cost of the building. The default solution under the code may have occasionally proven unfair, given the greater subjective value of a building for those who designed and built it, compared to the average market value or the subjective valuation of the unwilling owner of the land.
52. Horizontal partition may prevent valuable improvements, such as leveling the ground for agricultural or construction purposes, excavating for proper drainage, or more simply creating a well or a wine cellar and so on. For a historical survey of the evolving conceptions of physical unity in property, see also Bianca (1999).
53. Recently, common law courts have been relatively creative in figuring out ways to enforce contracts that create covenants designed to protect existing amenities in residential areas. Furthermore, legal systems will occasionally invent a new form of property. Despite these periodical innovations, this area of the law remains the most archaic. Rose (1999, pp. 213–14) observes that the common law system of estates in land now seems almost risibly crude and antiquated. As the author ironically points out, references to the 'fee tail' seldom fail to bring a smile.
54. Yiannopoulos (1983) notes the inadequacy of building and zoning ordinances to satisfy the needs of local property owners (for example, for the preservation of the subdivision style and so on). He also mentions that land developers have, since the turn of the century, imposed contractual restrictions limiting the use of property to enhance property values (for example, restricting use to certain specified purposes, prohibiting the erection of certain types of buildings, or specifying the material or the colors that may be used in the construction). Rudden (1987) observes, along similar lines, that although standard possessory interests involve exclusive and continuous possession, individuals may seek to acquire alternate interests such as a time-share, which is exclusive possession for repeated, short intervals. He thinks that servitude interests have seen the most innovation of late, and that security interests have seen the least innovation.
55. Real covenants and easements differ from one another in various respects. An affirmative real covenant is a promise to do an affirmative act (for example, a landowner's agreement with his/her homeowner's association to pay yearly fees, or a landowner's agreement to keep his/her lawn well trimmed). An affirmative easement, by contrast, is a right held by the owner of the benefiting land (the 'dominant estate') to use another party's land (the 'servient estate'). There is no affirmative obligation for the servient estate owner to do

anything. As generally explained, the servient owner only has a 'negative duty' to refrain from interfering with the other party's rights.

56. Under the modern contract doctrine, an assignee is liable for pre-existing contractual obligations only if he or she expressly assumes those obligations. The problem of the law of real covenants is thus principally concerned with situations where the assignee has not expressly agreed to assume the covenants of the previous owner. Enforcing these real covenants as mere contracts would often frustrate the goals pursued by the parties, given the frequent objective to allow the burdens and benefits of real covenants to pass to the successive owners of the underlying estates.
57. Yiannopoulos (1983) observes that the French Supreme Court (for example, Civ., 12 December 1899, D. 1900.1.361, with a note by Gény) recognized the effect against third parties of a property covenant relieving the operator of a mine from liability for damage to the surface. See also Bergel (1973).
58. See Rose's seminal 'Comedy of the commons' (1986), describing the origins of, and justifications for, common law doctrines and statutory strategies that vest collective property rights in the 'unorganized' public as a means of optimal resource management. Most recently, Henry Smith (2000) introduced the notion of semi-commons. These are property arrangements consisting of a mix of both common and private rights, with significant interactions between the two, observing that this property structure allows the optimizing of the scale of different uses of the property (for example, larger-scale grazing, smaller-scale grain growing and so on).
59. The idea of the anticommons in environmental regulation is explored further in Mahoney (2002).
60. Most recently, Dagan and Heller (2001) present the case of the liberal commons as a compelling illustration of efficient commons. Less obviously, we could imagine cases of purposely chosen anticommons. Examples of purposeful dysfunctional property fragmentation can be found in situations where unified property owners want to generate anticommons problems as a way of controlling the use of their property beyond the time of their ownership. An interesting real-life example is offered by the case of nature associations and mountain-hiking clubs that utilize anticommons-type fragmentation as a way to ensure long-term or perpetual conservation of the land in its current undeveloped state.

References

Austin, John (1832), *Lectures on Jurisprudence*, 5th edn, reprinted in R. Campbell (ed.) (1885), 2 vols, London: J. Murray.

Banner, S. (1999), 'Two properties, one land: law and space in 19th-century New Zealand', *Law and Social Inquiry*, (December), 807–52.

Bergel, Jean-Louis (1973), *Les Servitudes de Lotissement à Usage d'Habitation*, Paris: Librairie Générale de Droit et de Jurisprudence.

Bianca, Massimo C. (1999), *La Proprietà*, Diritto Civile, 6, Milan: Giuffrè.

Bouckaert, B. (1999), 'Original assignment of private property', *Encyclopedia of Law and Economics*, Cheltenham, UK, and Northampton, MA, USA: Edward Elgar, p. 1100.

Corbin, A.L. (1926), 'Assignment of contract rights', *University of Pennsylvania Law Review*, **74** (3), 207–34.

Dagan, H. and M.A. Heller (2001), 'The liberal commons', *Yale Law Journal*, **110** (1), 549–623.

Demsetz, H. (1967), 'Toward a theory of property rights', *American Economic Review*, **57** (2), 347–59.

Demsetz, Harold (1988), *Ownership Control and the Firm: The Organization of Economic Activity*, vol. I, Oxford: Basil Blackwell.

Dukeminier, Jesse and James E. Krier (2002), *Property*, 5th edn, New York: Aspen Law & Business.

Dwyer, John P. and Peter S. Menell (1998), *Property Law and Policy: A Comparative Institutional Perspective*, New York: Foundation Press.

Ellickson, Robert C., Carol M. Rose and Bruce A. Ackerman (eds) (1995), *Perspectives on Property Law*, 2nd edn, Boston, MA: Little, Brown.
Galgano, Francesco (1976), *Lex Mercatoria*, Bologna: Il Mulino.
Gambaro, Antonio (1995), *Il Diritto di Proprietà*, Milan: Giuffrè.
Hegel, Georg Wilhelm Friedrich ([1821] 1942), *Philosophy of Right*, trans. T.M. Knox, Oxford: Oxford University Press.
Kant, Immanuel ([1797] 1887), *Philosophy of Law*, trans. by W. Hastie, Edinburgh: T&T Clark.
LaFargue, Paul (1975), *The Evolution of Property: From Savagery to Civilization*, Chicago, IL: Charles H. Kerr.
Mahoney, J. (2002), 'Perpetual restrictions on land and the problem of the future', *Virginia Law Review*, **88** (4), 739–87.
Maine, Henry ([1861] 2000), *Ancient Law*, York, PA: Beard Group.
Mattei, Ugo (2000), *Basic Principles of Property Law*, Westport, CT: Greenwood Press.
Merrill, Th. and H. Smith (2000), 'Optimal standardization in the law of property: the *numerus clausus* principle', *Yale Law Journal*, **110** (1), 1–70.
Monateri, Piergiuseppe (1996), *Il Modello di Civil Law*, Turin: Giappichelli Editore.
Parisi, F. (1998), 'Customary law', in P. Newman (ed.), *New Palgrave Dictionary of Economics and the Law*, London: Macmillan Reference, pp. 572–9.
Parisi, F. (2002a), 'Entropy in property', *American Journal of Comparative Law*, **50**, 595–622.
Parisi, F. (2002b), 'Freedom of contract and the laws of entropy', *Supreme Court Economic Review*, **10**, 65–90.
Parisi, F. and E. O'Hara (1998), 'Conflict of laws', in P. Newman (ed.), *New Palgrave Dictionary of Economics and the Law*, London: Macmillan Reference, pp. 387–96.
Pipes, Richard (1991), *Property and Freedom*, New York: Alfred A. Knopf.
Posner, Richard A. (1998), *Economic Analysis of Law*, 5th edn, New York: Aspen Law & Business.
Rodotà, Stefano (1990), *Il Terribile Diritto*, 2nd edn, Bologna: Il Mulino.
Rose, C.M. (1985), 'Possession as the origin of property', *University of Chicago Law Review*, **52** (1), 73–88.
Rose, C.M. (1986), 'The comedy of the commons: custom, commerce, and inherently public property', *University of Chicago Law Review*, **53** (3), 711–81.
Rose, C.M. (1998), 'Canons of property talk, or, Blackstone's anxiety', *Yale Law Journal*, **108** (3), 601–32.
Rose, C.M. (1999), 'What government can do for property (and vice versa)', in Nicholas Mercuro and Warren Samuels (eds), *The Fundamental Interrelationship between Government and Property*, Greenwich, CT: JAI Press, pp. 209–22.
Rudden, B. (1987), 'Economic theory v. property law: the *numerus clausus* problem', in John Eekelaar and John Bell (eds), *Oxford Essays on Jurisprudence*, 3rd edn, New York and Oxford: Clarendon Press, pp. 239–63.
Schulz, N., F. Parisi and B. Depoorter (2002), 'Fragmentation in property: towards a general model', *Journal of Institutional and Theoretical Economics*, **158**, 594–613.
Smith, Adam (1776), *An Inquiry into the Nature and Causes of the Wealth of Nations*, reprinted in Robert Heilbroner (ed.) (1986), *The Essential Adam Smith*, New York and London: W.W. Norton.
Smith, H.E. (2000), 'Semicommon property rights and scattering in the open fields', *Journal of Legal Studies*, 29, 131–69.
Wenzel, M.A. (1993), 'Comment: the model surface use and mineral development accommodation act: easy easements for mining interests', *American University Law Review*, **42** (2), 607–81.
Yiannopoulos, Athanassios N. (1983), *Civil Law Property Coursebook: Louisiana Legislation, Jurisprudence, and Doctrine*, 3rd edn, Baton Rouge, LA: Claitor's.

4 Empirical issues in culture and property rights

*Seth W. Norton**

The paradox

There is ample evidence of a link between economic institutions and economic performance. Douglass North's seminal economic history of the West and allied analysis by Rosenberg and Birdzell document a relationship between economic institutions – property rights, the rule of law, and economic freedom and economic performance (North 1981, 1990, 1994; Rosenberg and Birdzell 1986). More recent research entails econometric analysis of the influence of economic institutions on economic growth and concludes that better specified or less attenuated property rights and the rule of law are associated with higher economic growth rates (Scully 1988, 1992; Barro 1991; Baumol 1994; Barro and Sala-i-Martin 1995; Knack and Keefer 1995; Knack 1996; Keefer and Knack 1997; Dawson 1998). More recent evidence shows that favorable economic institutions are particularly important for eliminating poverty (Scully 1997; Grubel 1998; Norton 1998, 2003).

The strong relationship between economic institutions and human well-being presents a paradox. The vast majority of the world's inhabitants would be better off living in regimes with well-specified and enforced property rights. Yet many of the world's inhabitants do not enjoy those benefits. Property rights and the rule of law are frequently weak and sometimes nearly non-existent. A profound question underlies the paradox. Why are well-specified and enforced property rights so rare?

Various explanations have been put forward to account for this puzzle. Demsetz (1964, 1967) and Anderson and Hill (1975) have developed what Eggertsson (1991) calls 'naïve' theories of property rights. In these views, property rights result from entrepreneurial incentives to establish and enforce rights. Relative benefits and costs create incentives to define and enforce property rights. Demsetz, and Anderson and Hill provide historical evidence consistent with that view. However, those accounts do not explain why the benefits and costs for institutional entrepreneurs vary so much across the world.

Some explanations for the failure of nation-states to adopt well-being enhancing institutions do exist. North (1990) identifies ideology and path dependency – the tendency for well-being inhibiting institutions to persist as

the source of less than ideal current institutions. His accounts are plausible but have a tautological flavor. Nation-states have their particular institutions because their property rights and legal systems reflect what their citizens (or at least their leaders) believe in (ideology) or what their ancestors believed in (path dependency).

In this chapter, I try to expand our understanding of the absence versus presence of property rights. I examine the hypothesis that cultural values determine the configuration of property rights in countries and hence determine at least in part the observable dispersion in economic performance across nations.

Demise of culture in economic analysis

Coase (1988, p. 3; [1991] 1994, p. 5) notes that modern economic analysis has progressively moved from close attention to facts to a focus on abstraction. Marshall (1920, p. 1) could describe economics as the study of human behavior 'in the ordinary business of life'. Many contemporary economists could do no such thing. The point applies to the culture/economics nexus as readily as any area. More generally, Huntington (2000) notes that in the 1940s and 1950s the study of culture was a widely and favorably viewed core concept in all of the social sciences. However, in the 1960s and 1970s, interest in culture 'declined dramatically' (2000, p. xiv). The pattern surely was evident in economics.

Formal economic analysis

Becker (1998) argues that economics largely ignores culture and does so at the expense of understanding important features of human behavior. There seems to be much support for Becker's point. Neoclassical economics puts heavy emphasis on maximizing behavior, equilibrium, the state of technology, and most of all relative prices. Culture gets short shrift and may even be treated derisively. For example, Ruttan (1991, p. 276) observes:

> Cultural considerations have been cast into the 'underworld' of developmental thought and practice. It would be hard to find a leading scholar in the field of developmental economics who would commit herself or himself in print to the proposition that in terms of explaining different patterns of political and economic development … a central variable is culture – the subjective attitudes, beliefs and values prevalent among dominant groups.

The point is particularly noteworthy in light of the fact that culture seems to be such a natural fit with economic development.[1]

The absence of any role for cultural explanations is somewhat ironic because there has been considerable attention given to the shortcomings of neoclassical economics, in particular the relative neglect of economic institu-

tions. In response, scholars have made great efforts in recent decades to integrate economic institutions with the modern analytical tools of economic theory.

Cultural relativism

The disdain for cultural explanations of economic phenomena has much broader roots than economic formalism. The concept of cultural relativism is a thoroughly entrenched precept of anthropology. Edgerton (1992) traces the nature and evolution of the idea. In the relativist perspective, evaluations of culture are inappropriate. The concept has foundations in Michel Eyquende Montaigne and David Hume but found its academic support in the American sociologist William Graham Sumner. However, the notion found wide dissemination in the field of anthropology as developed by Franz Boas at Columbia University, and his prominent students – Ruth Benedict, Melville Herskovits and Margaret Mead. Edgerton (1992, p. 26) notes that in more recent years, distinguished anthropologists such as Clifford Geertz and David M. Schneider, have expanded the concept to the point where cultures cannot even be compared.

One product of cultural relativism is the belief that comparing cultural variables is inappropriate in examining differences in groups of people, including economic differences. Thus, while economic formalism eschews culture as a determinant of economic behavior and performance, cultural relativism means that economic performance should not be used for comparative purposes. In more recent forms, this view has logically led to a world of epistemic skepticism where the concept of culture has nearly no meaning in the examination of social phenomena, including economic institutions.

Resurgence of culture in economic analysis

Despite the legacies of economic formalism and cultural relativism, culture as an analytic construct has emerged among a wide range of scholars in the last decade, especially those interested in the persistent differences in the wealth of nations. The proliferation of scholarship dealing with culture is so extensive that citations run the risk of woeful omissions. However, the role of various religions and ethnic issues has proved to be statistically robust in accounts of economic performance (Easterly and Levine 1997; Knack and Keefer 1997; Lazear 1999; Tornell and Lane 1999; Norton 2000), and in variations in economic institutions (La Porta et al. 1997, 1999; Norton 2000).

Failure of convergence

One reason for renewed interest in culture stems from the failure of growth rates of rich and poor countries of the world to converge, as predicted by neoclassical growth theory.[2] The important point is that simple neoclassical

models have been found to be inadequate. The economics profession has responded in two directions. First, the underpinnings of neoclassical growth models have been revised with 'endogenous growth models' that seem to fit the facts of limited convergence (Romer 1986). Second, with more specialization within the economics profession, the neo-institutional approach has led to comprehensive and systematic probes of why the whole world is not developed. The centrality of economic institutions has emerged as the foundation for growth.

Institutional links

The neo-institutional approach to economics includes a wide variety of subjects and methods. Examples range from formal theory as in Grossman and Hart (1986) on vertical integration to evolutionary game theory (Hirshleifer 1987), or to case studies of solutions to common-pool resource allocation (Libecap 1990; Ostrom et al. 1994). Two approaches that are particularly relevant for this chapter are long-run historical studies such as North (1981, 1990) and Rosenberg and Birdzell (1986) and cross-sectional econometric studies of economic well-being such as Scully (1988), Barro and Sala-i-Martin (1995) and Knack and Keefer (1995). In both approaches, the evidence suggests that economic well-being is enhanced by economic institutions that specify and enforce property rights.

The most relevant recent development in the economics of institutions has been the use of cultural variables – particularly ethnicity and religion, to examine why some countries have welfare-enhancing institutions and other countries do not (La Porta et al. 1997, 1999; Norton 2000).

Entrepreneurship and the culture of capitalism

Besides the neo-institutional response to failure of convergence, culture has emerged in its own right as a foundation for the observed phenomenon that not all groups of people are the same with respect to the perception of profit opportunities and the appropriate strategies of coping with uncertainty (Berger 1991). Some cultures – including various minority groups within larger dominant cultures – are more adept at generating wealth than others. Examples include studies by Greif (1989, 1994) and Landa (1994).

A related approach is the role of culture in determining the prerequisite values of modern capitalist societies – education, a strong work ethic, a cognitive link between prosperity and productivity, and a host of others. For example, Grondona (2000) identifies no less than 20 contrasting cultural factors that affect economic performance, and in turn, human well-being.

The systematic study of the cultural foundations of capitalism also rests on the failure of basic economic predictions. Two events are noteworthy. First, a widely held view in the mid-twentieth century was that once the

colonies lost their imperial shackles, their economies would flourish. Second, the demise of Soviet socialism was presumed to permit the natural capitalists of the former Soviet countries to exercise their freedom and in turn prosper. In both events, prosperity has been elusive and even nonexistent. Explanations for the failure of prosperity to emerge have led to a focus on culture.[3]

Culture and institutions

While our understanding of economic behavior is enhanced by the failure of economic predictions, the connection between culture and institutions is unclear. The study of institutions and culture suggests that both forces are important and crucial in enhancing well-being, but the details of the relationship are somewhat clouded.

Consider P.T. Bauer (1971, 1984) and Douglass North (1981, 1990, 1994), two of the foremost scholars in moving economics toward a richer understanding of the role of transaction cost-reducing institutions and, in turn, enhancing trade and improved economic performance. Lavoie and Chamlee-Wright (2000) note that both Bauer and North strongly affirm the role of cultural foundations of economic institutions, but a full accounting of the culture/institutions nexus is not central to their analysis. Other studies (for example, Abrams and Lewis 1995; Hanson 1999; La Porta et al. 1999; Mahoney 2001) do use cultural variables such as a religion to 'explain' variations in economic behavior and institutions. However, their analysis is characteristically empirical with minimal links to the anthropological, economic, or historical literature on culture.

Culture as an ambiguous concept

A striking feature of the study of culture is that the meaning of the term 'culture' is contested. Many definitions of culture exist. Berger (1991) notes that two themes are evident. One set of definitions assumes a 'restricted view'. In this perspective, culture entails largely symbolic dimensions. A second set of definitions assumes an 'inclusive view'. In this perspective, culture is a way a life. Similarly, Lal (1998) uses a cultural dichotomy, where material beliefs dealing with 'making a living' are distinct from cosmological beliefs dealing with issues of 'purpose, meaning and relationship to others'. Lal provides a sweeping history of economic life relying on both types of culture, but with special emphasis on the role of cosmological beliefs in generating the 'Promethean growth of the West' (1998, p. 8). Accordingly, it warrants special attention in any discussion of the cultural foundations of economic institutions.

Definition

Given the ambiguity of the term 'culture', some flexibility is warranted. Lal's (1998) 'cosmological beliefs' foundation of culture seems to generate a coherent account of a broad range of people groups. Similarly, Greif's (1994) analysis of Genoese merchant traders and Maghribi traders stresses 'cultural beliefs'. That approach and Lal's (1998, p. 8) emphasis on 'purpose, meaning and relationship' seem congruent with Hofstede's (1991) view of culture as 'mental programming' – the transmitted values of a community with particular emphasis on their meaning, purpose and relationships and ultimately on their beliefs and values – the operative meaning of culture in this chapter.

Cultural comparative advantage

Treating culture as largely a system of values and beliefs and arguing that economic institutions reflect those values and beliefs is consistent with the detailed micro-historical analysis of Greif (1994) and the sweeping macro analysis of Lal (1998). However, two additional features of cultural analysis must temper that approach.

Peter Berger (Berger and Hsiao 1988) argues that the economic concept of comparative advantage applies to culture as well as to natural resources and other endowments. Thus, countries should find some advantages to using features of their cultures in creating the benefits from trade. Lavoie and Chamlee-Wright (2000) amplify the concept by noting that a productive 'spirit of enterprise' can exist within differing cultural contexts.

Cultural nationalism

A second feature that warrants a special disclaimer is the concept of cultural nationalism. Many nation-states are in fact multicultural. Values and beliefs vary widely across many countries, especially geographically large ones. Consequently, empirical analysis of any sort encounters the potential pitfalls of overgeneralization.

Measures of culture

If cultures are presumed to affect the basic economic institutions of large groups of humanity, then some meaningful basis for examining cultures is warranted. However, the dominant tradition of anthropology has been ethnographic field work. That approach tends to stress the nuances of people groups – generally at less than nation-state level. The practice of focusing on the unique cultural features renders comparative analysis difficult. Accordingly, an examination of cultural foundations requires contemporaneous, normalized systematic data across nation-states.

Hofstede data

Geert Hofstede (1991) has compiled a truly remarkable data set for examining the effects of cultural values and beliefs on economic life. During the years 1967–73, the IBM corporation conducted 117 000 employee attitude surveys for its overwhelmingly native workforce in 67 countries. The surveys entailed 60 core questions and 66 recommended questions. The focus of the questions was personal goals and beliefs – including work goals, perceptions and satisfactions. The questions were originally developed in English but later translated into native languages of the respective countries. Most questions used a five-point (quasi-interval) scale between two poles. For example, respondents were requested to identify their answer between 'very poor to very good', 'utmost importance to no importance', and 'strongly agree to strongly disagree'.

The survey responses were recorded along with substantial demographic and occupational data. The data were standardized in terms of deviations from the mean. Hofstede (1991) examined the data via cross-tabulations and correlations. Using factor analysis, he was able to identify patterns associated with occupation, sex and country in terms of values and beliefs. The patterns for IBM employees in each country were scored and used for comparisons regarding some salient features of employee's values and beliefs. Four categories of values and beliefs – individualism, masculinity, power and uncertainty – were identified.[4]

Individualism

Hofstede identifies individualism as a cultural dimension that reflects people's identity as individuals or as members of a group (collectivism). The concept and its contrast are well established in the anthropological literature. Individualistic cultures are ones that predominantly consist of people who view themselves as responsible for themselves and their immediate families, while collectivist cultures are ones that predominantly reflect people who view themselves as protected by their membership in larger groups – communities, religions, extended families, and even social classes.

The measure of individualism versus collectivism is a continuum derived from a series of questions that stress 14 work goals of IBM employees. The questions deal with the value of using one's skills on the job or having good working conditions (Hofstede 1982, p. 155). Countries with a high proportion of individualistic employees tend to value personal time, freedom on the job, and sense of accomplishment from the job. Countries with a high proportion of collectivist employees tend to value training, good physical conditions at work, and the use of skills on the job.

Masculinity
Masculinity is a second cultural dimension developed by Hofstede. The measure identifies people who favor action and achievement as opposed to affiliation and acceptance of failure, which are linked to femininity.

The masculinity measure is also based on work goal questions. Because considerable segregation by sex existed in the IBM employee sample, the data used for this measure are restricted to occupations that employed both women and men in large proportions. Countries are characterized as more masculine where employees tend to view their jobs as more central to their life and where employees tend to value earnings, advancement and challenge. Countries tend to be categorized as feminine where employees view their jobs as less central to life and where a friendly atmosphere, position security and cooperation are valued.

Power distance index
Power distance reflects people's belief that power and status are unequal and that this inequality is acceptable. This dimension deals with values and beliefs that are nearly universal (Hofstede 1982, p. 65) but with considerable variation across groups of people. Hofstede uses the concept as closely linked to superior and subordinate relationships and efforts by parties in those relationships to maintain or change the nature of the relationship.

Hofstede's categorization rests on questions to employees (managers were excluded from the sample for this measure). The questions referred to the type of managers employees preferred and what type of managers they actually had. Both questions focused on four characteristics – autocratic, persuasive/paternalistic, consultative and democratic. For example, employees were asked (Hofstede 1982, p. 73): 'How frequently in your experience, does the following problem occur: employees being able to express disagreement with their managers?'. Respondents were instructed to indicate their experience on a five-point scale between 'very frequently' to 'very seldom'. Countries were measured as more hierarchical – a high 'power distance index' in Hofstede's terms, if the employees preferred and had managers with more autocratic styles, lower personal trust and greater privilege than in countries with lower measures on the power distance index.

Uncertainty avoidance
Uncertainty avoidance is a cultural dimension where people eschew ambiguity. A predominance of uncertainty avoidance in a culture should lead to reliance on contingencies, plans and specified routines as well as rule of thumb that restrict human actions. Cyert and March (1963) observed that people commonly avoid uncertainty by emphasizing rules that (i) stress the

short run over the long run and (ii) avoid plans that relate to the future and the need to anticipate the future.

Hofstede builds on those observations by categorizing countries with respect to the proclivities of IBM employees in a particular country to answer questions indicating uncertainty avoidance. The questions deal with employees' willingness to concentrate on rule orientation, their employment stability and the incidence of job stress. Rule orientation deals with employees' willingness to violate company rules when the context indicates that the firm would benefit from not following the rules. Employment stability deals with employees' expected length of employment with the firm. Stress deals with the frequency that employees are nervous or tense. Employees with greater uncertainty avoidance are less willing to break company rules, have a longer expectation of staying with the firm, and have more frequent stress on the job compared to other employees. Countries with greater proportions of such workers have higher measures of the uncertainty avoidance scale.

Descriptive statistics

Descriptive statistics of Hofstede's four cultural measures are shown in Table 4.1.

Table 4.1 Varieties of cultural values

Cultural value	Highest	Middle	Lowest	Mean	Std dev.
Individualism	United States (91)	Iran (41)	Guatamala (6)	41.1	23.7
Masculinity	Japan (95)	Singapore (48)	Sweden (5)	48.7	16.8
Power	Malaysia (104)	Greece (60)	Austria (11)	60.0	21.2
Uncertainty avoidance	Greece (112)	Germany (65)	Singapore (8)	64.7	22.4

Note: N = 64. Data are from Hofstede (1991). The numbers in parentheses are the Hofstede measure for that country and value.

The data show the mean and standard deviations of the cultural value measures and examples of the highest and lowest countries as well as an example of a middle country – the one closest to the sample mean. The data in Table

4.1 show considerable dispersion of the cultural values, indicating a reasonable basis to examine the culture/institutions nexus.

Testable hypotheses

Hofstede's data provide a straightforward means for testable hypotheses regarding culture and economic institutions. Greif (1994) and Lal (1998) build a compelling case that the prerequisites of a market economy include belief systems that rest on individualism and not collectivism. Moreover, Greif argues and Lal affirms that an individualist culture requires a formal legal system to facilitate exchange. The meaning of individualism and the individualism – private property nexus is described by Macfarlane (1987). He points to the existence of autonomous, equal (before the law), separate individuals whose interest dominates any constituent group within the polity. Individualism is reflected in the idea of individual property and supporting liberty of the individual. Accordingly, it is most reasonable to hypothesize that nation-states with well-specified property rights and the rule of law rank higher on Hofstede's individualist versus collectivist measure. Note that the property rights and the rule of law are positively related ($r = 0.89$). Hence, they appear to be strong complements.

Clear hypotheses regarding Hofstede's other measures are more difficult to generate. However, some additional, albeit more tentative, hypotheses seem fitting. Macfarlane notes that the culture in England led to an extraordinary focus on equality before the law: 'the law enshrined an obsession with property, which was conceived of as virtually private, rather than communal' (Mcfarlane 1987, p. 145). The England that gave rise to such veneration of the law and private property was also unique in its absence of inherent legal authority of men over women and almost remarkably non-hierarchical by comparative standards. To the extent that law and property rights are linked with these cultural features, we might expect that their occurrence elsewhere would lead to similar results. Accordingly, we can test the hypothesis that Hofstede's measures of hierarchy and masculinity are negatively related to well-specified property rights and the rule of law.

In contrast to the individualism versus collectivism, masculinity and power, Hofstede's uncertainty avoidance measure is problematic. Markets reduce uncertainty by facilitating the transformation of uncertainty into risk. Thus, the underlying institutions of capitalism, property rights and the rule of law, may be more common where people do not exhibit avoidance of uncertainty. Moreover, the cognitive present mindedness of uncertainty avoidance would seem likely to retard investment in an institutional infrastructure with a focus on the future – the practical effects of property rights and the rule of law. Hence, we can hypothesize a negative relationship between uncertainty avoidance and property rights and the rule of law. However, it is also likely that

these institutions reduce uncertainty and may lead to less uncertainty avoidance. Hence our interpretation of uncertainty avoidance is ambiguous.

Measuring property rights

In recent years, empirical research on the role of economic institutions has relied on a variety of measures (Scully 1988, 1992, 1997; Barro and Sala-i-Martin 1995; Knack and Keefer 1995; Knack 1996; La Porta et al. 1999). Several of these are used in the analysis below.

Three measures derived from the *International Country Risk Guide* (ICRG) published by Political Risk Services are used in the analysis in this chapter. The numbers are ratings of a broad sample of countries on several dimensions. Data are available as early as 1982. The data used in the estimate are from the first available year, usually 1982, through 1995. The data are described in detail in Barro and Sala-i-Martin (1995) and Knack and Keefer (1995).

Although multiple dimensions of property rights exist, for simplicity and tractability, property rights are presumed to entail the absence of 'takings' and the sanctity of contract. Two ICRG measures provide a good fit with these concepts. The first measure is expropriation risk where increasing scores indicate a lower probability that private property will be confiscated by the government. The second measure is repudiation of contracts. Increasing scores indicate a lower probability that the government will renege on contracts. These two measures are combined with equal weight and normalized between zero and one, to constitute the property rights measure in the empirical tests. Empirical evidence by Norton (2003) shows that this measure has a strong explanatory power in poverty reduction and enhancing economic development.

A third ICRG measure is entitled the rule of law. Increasing scores indicate orderly transitions of power and authoritative adjudication of disputes. Knack and Keefer (1995, p. 225) describe the measures in terms of accepting 'established institutions to make and implement laws and adjudicate disputes'. The rule of law means that citizens rely on established legal procedure and orderly successions to authoritative positions and avoid the use of physical force or illegal means to settle claims. Empirical studies by Knack and Keefer (1995) and Barro and Sala-i-Martin (1995) show that the rule of law is a particularly potent institutional determinant of economic growth.[5]

Descriptive statistics for the property rights are shown in Table 4.2. For this sample, South Africa is in the middle of the distribution for the property rights measure, while Iraq has the lowest score and Switzerland has the highest value. For the rule of law measure, Venezuela is in the middle of the distribution, while Colombia has the lowest measure, and Australia and nine other countries have the highest measure for the rule of law.

Table 4.2 Descriptive statistics: property rights and the rule of law

Measures	Mean	Std dev.	Minimum	Maximum
Property rights	0.75	0.18	0.26	1.00
Example	South Africa (0.75)		Iraq	Switzerland
Rule of law	0.64	0.25	0.23	1.00
Example	Venezuela (0.68)		Colombia	Australia*

Note: *Also Austria, Canada, Denmark, Finland, the Netherlands, Norway, Sweden, Switzerland and the USA.

Empirical tests

To examine the four testable hypotheses regarding the link between the Hofstede measures of cultural characteristics and the property rights measures, I estimate the following equations:

$$PR_i = \beta_{01} + \beta_1 IND_i + \beta_2 MASC_i + \beta_3 POWER_i + \beta_4 UNCER_i + u_{1i} \quad (4.1)$$

$$ROL_i = \beta_{02} + \beta_5 IND_i + \beta_6 MASC_i + \beta_7 POWER_i + \beta_8 UNCER_i + u_{2i} \quad (4.2)$$

where PR and ROL represent the property rights and rule of law measures, and IND, MASC, POWER and UNCER represent Hofstede's measures on individualism, masculinity, power distance and uncertainty avoidance.

Estimates of equations (4.1) and (4.2) are shown in Table 4.3. The estimates are shown for the full sample of countries and a smaller restricted sample with less measurement error.[6] The data in the table document a powerful positive and powerful relationship between property rights and more individualist cultures and a similar and statistically even stronger relationship between the rule of law and individualist culture. In both cases, the relationship is approximately the same for the full and restricted samples.

The data also document a negative relationship between Hofstede's power index and property rights and the rule of law. The relationship is less robust than the individualist measure, and it deteriorates in the restricted sample. On the other hand, the masculine and uncertainty avoidance measures yield no explanatory power. Indeed, estimating equations (4.1) and (4.2) with only the individualistic and power measures puts the adjusted R-squared measure at

Table 4.3 Cultural values, property rights and the rule of law

Dependent variable	Intercept	Individualist	Masculine	Power	Uncertainty avoidance	Adj. R^2	S.E.R.*
(Sample)							
Property rights (full: N = 64)	0.780 (7.19)	0.310 (3.34)	0.468 (0.49)	–0.305 (–3.00)	0.000 (0.11)	0.458	0.132
Property rights (restricted: N = 50)	0.768 (7.53)	0.323 (3.81)	0.297 (0.36)	–0.157 (–1.63)	–0.060 (–0.91)	0.476	0.115
Rule of law (full: N = 64)	0.752 (4.21)	0.690 (4.54)	–0.240 (–1.31)	–0.420 (–2.45)	–0.020 (–0.12)	0.461	0.184
Rule of law (restricted: N = 50)	0.789 (3.71)	0.690 (3.96)	–0.250 (–1.24)	–0.260 (–1.31)	–0.140 (–0.92)	0.429	0.191

Note: The full sample includes seven Arab, four East African and three West African country groups. The aggregate values for the three country groups are imputed to each separate member of the group in the full sample. The restricted sample excludes the three country groups.

* Standard Error of Regression.

about 50 per cent for the property rights measure and about 45 per cent for the rule of law measures. Thus, the two cultural measures seem to 'explain' roughly half of the variation in property rights and the rule of law. The results are consistent with the accounts by Macfarlane, Greif and Lal. Other measures of cultural values appear irrelevant.

Additional variables

Besides values and beliefs, other cultural variables are likely to affect the economic institutions in a country. Ethnic and linguistic homogeneity or fractionalization and religion are clear candidates. There are sound theoretical foundations for the proposition that ethnic and linguistic differences reduce the cooperative behavior of human aggregations (Tornell and Lane 1999). Moreover, there is solid empirical support that such reduced cooperation leads to weaker property rights systems and retarded economic performance (Easterly and Levine 1997; Knack and Kiefer 1997; Norton 2000). Thus, cultural factors beyond basic values and beliefs may also play a role in determining property rights systems. Of particular relevance is the question whether or not diversity measures weaken or complement the statistical relationships documented in Table 4.3.

The results from previous empirical research suggest that property rights are likely to be better protected and the rule of law is likely to be better established as language or ethnic homogeneity increase and the opposite for ethnic fractionalization.[7] Three measures of ethnic diversity/homogeneity are readily available for analysis – the size of the largest language group in a country as a proportion of the population (Muller 1964), the proportion of the largest ethnolinguistic group in a country (Barrett 2001), and a widely used measure of ethnic fractionalization (La Porta et al. 1997).[8]

Previous research has also shown that the world's major religions affect the institutional infrastructure of nation-states (La Porta et al. 1997, 1999). The rationalization is somewhat *ad hoc*. There are appeals to the relative hierarchical nature of religions or the governmental 'interventionist' tendencies of religions and related attenuation of property rights and the rule of law. Protestantism is presumed to enhance property rights and the rule of law while Catholicism and Islam are presumed to retard them (La Porta et al.1997).

The most relevant consideration is the effect of the inclusion of religion on the cultural values measures. If the inclusion of religion weakens the Hofstede measures, that is, if, using the language of econometrics, the estimated coefficients prove 'fragile' to the inclusion of religion variables, then there is a reasonable basis to infer that religion affects institutions through values and beliefs, but other features of religion are also important. Potential factors include traditions of hierarchical organization or religious doctrine that lead to attenuation of property rights or weakening the rule of law. On the other

hand, if religion variables are significant, and Hofstede's measures also remain significant, then a reasonable inference is that religion has effects primarily beyond measured cultural values and beliefs.

Estimates of equations (4.1) and (4.2) with the diversity and religion variables included are shown in Table 4.4. The ethnic and language variable estimates are statistically significant with both language group size and ethnic fractionalization strongly related to both property rights and the rule of law. The size of the largest ethnolinguistic group has the predicted sign in both cases but is only marginally significant.

The role of religion is less uniform. Protestantism plays no discernible role. The Catholic proportion is only marginally significant in the property rights case and not significant at all for the rule of law measure. The Muslim proportion of a nation's population is strongly negative for the property rights measure, and the relationship is robust. On the other hand, the Muslim proportion is at most marginally significant for the rule of law measure.

The increased explanatory power attributable to ethnicity and religion is noteworthy. The adjusted R^2 values increase from the 43 to 47 per cent range to the 52 to 63 per cent range. Thus, the data indicate that ethnicity and religion help explain the variation in property rights and the rule of law beyond just the effects of values and beliefs that are captured in the Hofstede measures.

The most relevant feature for Table 4.4 is the effect on the Hofstede measures when ethnicity and religion are included. The masculinity and uncertainty avoidance measures are unchanged in these estimates; they remain irrelevant. The power measure loses its significance, perhaps indicating that the power measure was simply capturing some features of ethnic diversity or the impact of hierarchical religions. More importantly, individualism retains its statistical explanatory power, and the size of the estimated coefficients is approximately the same. In short, the value of an individualist ethos on property rights and the rule of law is palpable even when religions and the debilitating effects of ethnic diversity are considered.

An additional set of variables that merit examination is legal origin. Douglass North (1990) rationalized the failure of economic systems to evolve toward growth-friendly configurations by invoking the concept of 'lock-in' or path dependency. In this framework, history is determinative. Path dependency has been used to rationalize the adherence to existing standards for multiple economic agents because the costs of switching are prohibitive. For example, when standard railroad gauges and equipment became accepted, it became costly for any railroad to use any other size. If a railroad did switch gauges and equipment, rail cars from its lines could not be used on the lines with the standard gauge – severely limiting their use for interregional shipping. Similar examples are thought to exist.

Table 4.4 Determinants of property rights and the rule of law: cultural values, ethnicity and religion

Independent variable	Coefficient (*z*-statistic)					
	Property rights			Rule of law		
Intercept	0.64 (5.40)	0.69 (5.81)	0.84 (7.61)	0.50 (2.48)	0.49 (2.32)	0.81 (3.38)
Individualistic	0.33 (3.90)	0.35 (4.20)	0.33 (4.50)	0.55 (3.33)	0.68 (4.66)	0.62 (3.83)
Masculine	0.00 (0.04)	0.03 (0.30)	–0.03 (–0.29)	–0.19 (–0.98)	–0.14 (–0.71)	–0.24 (–1.15)
Power	–0.09 (–0.87)	–0.11 (–1.10)	–0.08 (–0.90)	–0.16 (–0.94)	–0.14 (–0.80)	–0.17 (–0.94)
Uncertainty avoidance	–0.05 (–0.60)	–0.04 (–0.43)	0.00 (0.46)	–0.18 (–1.09)	–0.10 (–0.05)	–0.09 (–0.50)
Language group	0.19 (2.95)	–	–	0.39 (3.25)	–	–
Ethnolinguistic group		0.11 (1.66)	–	–	0.23 (1.78)	–
Ethnic fractionalization		–	–0.17 (–3.23)	–	–	–0.23 (–1.94)
Catholic	–0.09 (–1.78)	–0.06 (–1.27)	–0.10 (–2.26)	–0.05 (–0.63)	–0.01 (–0.13)	–0.08 (–0.87)
Muslim	–0.28 (–4.46)	–0.22 (–3.79)	–0.22 (–3.48)	–0.20 (–1.88)	–0.13 (–1.36)	–0.19 (–1.74)
Protestant	–0.06 (–0.63)	–0.05 (–0.56)	–0.08 (–0.93)	0.14 (0.73)	0.15 (0.74)	0.12 (0.57)
Adj. R^2	0.619	0.577	0.635	0.553	0.518	0.531
S.E.R.*	0.112	0.117	0.102	0.169	0.174	0.173
N	60	63	58	60	63	58

Note: All coefficients except the intercept and diversity measures should be multiplied by 0.01.

* Standard Error of Regression.

North's contention is that there are similar path dependencies regarding economic institutions. Given market imperfections attributable to transaction costs and increasing returns in the creation of institutions, comparatively poor economic institutions can not only exist, but also persist. In short, once dispersed throughout a polity, institutional standards acquire a life of their own.

To examine the effects of path dependency, I use the categorical variables developed by La Porta et al. (1999) as well as the religion and ethnicity variables, to estimate further augmented versions of equations (4.1) and (4.2). The categorical variables (0 or 1) represent legal origins identified as English and German or Scandinavian.[9] The results for the estimates are reported in Table 4.5. The data show differential effects for the legal origin coefficients. The English or French origin countries have weaker property rights and rule of law than the German or Scandinavian ones. At least by these measures, there seems to be something to the path-dependence account.

The data show some different results. Hofstede's masculinity and uncertainty measures are negative and significant, as are the Protestant coefficients. The first two are somewhat hard to interpret, but the Protestant effect presumably reflects mulicollinearity. The masculinity effect is consistent with Macfarlane's observation regarding the role of greater gender equality in regimes with stronger protection of private property and commitment to the rule of law. However, it is certainly possible that these effects simply reflect correlation with the legal origin variables. Also, the ethnicity variables lose some explanatory power when the legal origin variables are included, although the language group size and ethnic fractionalization remain significant for the property rights measure.

The most significant result is the robustness of the individualism coefficients documented in Table 4.5. The estimated variables are not fragile to the inclusion of all other variables. The most evident conclusion is that there is a strong relationship between property rights, the rule of law, and cultural values and beliefs that affirm individualism. Thus, the data are consistent with historical accounts by Macfarlane, Greif and Lal. Culture is important to economic institutions. Individualist as opposed to collectivist cultures promote property rights and the rule of law.

Interpretation

The data presented above indicate that economic institutions are contextual. Current economic institutions reflect the distant past. Current economic institutions at least to some degree reflect the inherent conflicts and cooperative behavior in a polity. Most importantly, the values and beliefs of citizens of nation-states are related to the economic institutions. This observation constitutes an important foundation for understanding crucial determinants of economic well-being.

Table 4.5 Determinants of property rights and the rule of law: culture and path dependency

Independent variable	Coefficient (*z*-statistic)					
	Property rights			Rule of law		
Intercept	0.74 (5.78)	0.72 (5.88)	0.82 (7.82)	0.57 (2.31)	0.59 (2.73)	0.78 (3.54)
English	–0.11 (–0.26)	0.04 (0.97)	0.03 (0.70)	–0.03 (–0.37)	0.06 (0.82)	0.02 (0.27)
German/ Scandinavian	0.32 (4.67)	0.28 (4.70)	0.26 (5.00)	0.42 (3.35)	0.42 (4.05)	0.4 (3.80)
Individualistic	0.49 (6.59)	0.47 (6.46)	0.45 (7.04)	0.87 (6.17)	0.83 (6.38)	0.8 (6.04)
Masculine	–0.29 (–2.97)	–0.24 (–2.41)	–0.24 (–2.76)	–0.58 (–3.05)	–0.55 (–3.02)	–0.56 (–3.06)
Power	0.11 (1.16)	0.3 (0.33)	0.03 (0.41)	0.23 (1.28)	0.05 (0.31)	0.02 (0.12)
Uncertainty avoidance	–0.18 (–2.15)	–0.09 (–1.13)	–0.12 (–1.73)	–0.27 (–1.69)	–0.22 (–1.56)	–0.24 (–1.71)
Language group	0.1 (1.78)	–	–	0.09 (0.89)	–	–
Ethnolinguistic group	–	0.05 (0.81)	–	–	0.14 (1.25)	–
Ethnic fractionalization	–	–	–0.12 (–2.47)	–	–	–0.14 (–1.41)
Catholic	–0.01 (–0.11)	0 (0.70)	–0.02 (–0.39)	0.07 (0.72)	0.11 (1.23)	0.07 (0.78)
Muslim	–0.24 (–3.85)	–0.22 (–3.87)	–0.18 (–3.27)	–0.18 (–1.50)	–0.11 (–1.12)	–0.12 (–1.04)
Protestant	–0.5 (–4.19)	–0.4 (–3.73)	–0.39 (–4.22)	–0.4 (–1.70)	–0.37 (–1.82)	–0.35 (–1.67)
Adj. R^2	0.698	0.687	0.745	0.58	0.621	0.63
S.E.R.*	0.095	0.1	0.085	0.158	0.155	0.154
N	51	62	58	51	62	58

Note: All coefficients except the intercept, diversity and legal origins measures should be multiplied by 0.01.
* Standard Error of Regression.

Culturalists versus institutionalists

The results documented above shed some light on an important debate regarding culture and institutions (Lavoie and Chamlee-Wright 2000). Two contrasting views exist. The 'institutionalist' view holds that economic institutions are the key to economic well-being, and hence, nation-states can export superior institutions to other nation-states, while the 'culturalist' view holds that economic well-being is largely culturally determined and less exportable. The strong link between property rights, the rule of law, and individualistic values and beliefs raises skepticism about the exportability of institutions. Thus, despite the dangers of overgeneralization, property rights and the rule of law seem to be more at home in some cultures than in others. Exploring the robustness of the point and implications of that constitutes a major research agenda.

Dark side of individualism

The last decades of the twentieth century saw the collapse of militarily powerful regimes that decried the institution of private property and the rule of law. In an ironic twist of Marxist doctrine, it seems that it was well-being regarding institutions that 'withered away'. However, the failure of well-being-enhancing institutions to emerge in those countries underscores the futility of imposing growth-friendly institutions by good intentions or through simple natural selection.

Much more importantly, the same period saw an acceleration of cultural trends in the West that undermine the cultural beliefs underlying the market economies of the West. Lal (1998) documents cultural attacks on civil society in the West and market-friendly institutions in particular, including a parasitic welfare state that erodes the foundations of individualism. Thus, it is plausible that the cultural foundations that led to the Promethean growth of the West may lead to destructive forms of individualism destroying institutions that generated the Promethean growth. The message in the data completely ignores this not so sanguine speculation.[10]

Conclusion

Property rights and the rule of law enhance human well-being, yet they are absent in varying degrees across much of the world. The pattern is puzzling and understanding this condition would help us understand one of the great issues of economic life. Why are such beneficial human arrangements not more common?

The data presented in this chapter support the view that they are not more common because nation-states seem to be locked in to some long-standing institutional endowments. These conditions are affected by religion and ethnic diversity. However, the most striking fact is that they are not more common

because cultural values and beliefs seem to profoundly affect the institutional configuration. The data are only suggestive, but they do indicate that simply reforming institutions may require a lot more than the imposition of a set of desirable legal arrangements. Institutional reform towards growth-friendly arrangements may well depend on underlying changes in values and beliefs. The results are sufficiently strong to warrant more detailed analysis of the culture/institutional nexus and the limits and prospects for institutional reform.

Notes

* Financial support from the Earhart Foundation is gratefully acknowledged. I have received helpful comments and suggestions from P.J. Hill. I am responsible for any remaining errors.

1. A popular textbook in development economics, Todaro (1997) does not contain the term 'culture' in the subject index.
2. The reference is to gross or unconditional convergence, especially between the rich and poor nations of the world. Convergence does occur when other variables such as human capital levels, investment and political instability are included in the analysis. See Baumol (1994) and Barro and Sala-i-Martin (1995).
3. The observations are detailed in Harrison (2000) and Huntington (2000).
4. Hofstede (1991) developed a fifth cultural dimension, long-term orientation. The data were collected after the initial compilation and thus may not be congruent with the other measures. Moreover, the long-term orientation data are not available for the full sample of countries and hence reduce the sample size for empirical tests by more than one-half. Preliminary estimates that included long-term orientation indicate that it is not statistically significant at conventional levels.
5. The property rights and rule of law measures are rescaled from the original ICRG ranking to values between zero and one. Accordingly, the measures are potentially censored, so Tobit regression estimates are used rather than ordinary least squares.
6. The full sample includes several countries that were consolidated as Arab-speaking countries, including Egypt, Iraq, Kuwait, Lebanon, Libya, Saudi Arabia and the United Arab Emirates; East African countries, including Ethiopia, Kenya, Tanzania and Zambia; and West African countries, including Ghana, Nigeria and Sierra Leone. The full sample includes these countries with Hofstede values for the group of countries imputed to the individual countries.
7. The point may apply more to property rights where government action is often inimical to the full use and alienation of property rights. The application to the rule of law is less obvious because interventionist governments might still affirm the political stability inherent in the ICRG's rule of law measure, which in the original formulation is described as 'law and order tradition'.
8. The ethnic variables are described in Easterly and Levine (1997) and Knack and Keefer (1997) and Norton (2000). The data are from Muller (1964), La Porta et al. (1999) and Barrett (2001).
9. The French effect is captured by the intercept term. Preliminary regression analysis indicated that while the Scandinavian coefficients uniformly exceed the German coefficients, the difference is not statistically significant. Latitude is included because geography may play a role and La Porta et al. (1999) found some statistically significant effects for latitude as a determinant on property rights and other measures of government quality.
10. Lal notes that some of the debilitating effects of individualism are less pronounced in other countries. Japan is one case. It is customary to treat Japan as communal or collectivist (for example, Temin 1997), and hence, its resilience to destructive cultural forces and the debilitating effects of the welfare state (for example, divorce and single-parent families),

means that a more communal culture might be somewhat more hospitable to market-friendly institutions in the future than the individualistic West. However, it is noteworthy that Japan is above the mean on the individualist measure, ranking 22 out of 64 sample countries used in this study.

References

Abrams, B.A. and K.A. Lewis (1995), 'Cultural and institutional determinants of economic growth: a cross-section analysis', *Public Choice*, **83**, 273–89.

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.

Barrett, David. B. (2001), *World Christian Encyclopedia*, New York: Oxford University Press.

Barro, R.J. (1991), 'Economic growth in a cross section of countries', *Quarterly Journal of Economics*, **106**, 407–43.

Barro, Robert J. and Xavier Sala-i-Martin (1995), *Economic Growth*, New York: McGraw-Hill.

Bauer, Peter T. (1971), *Dissent on Development*, Cambridge, MA: Harvard University Press.

Bauer, Peter T. (1984), *Reality and Rhetoric: Studies in the Economics of Development*, London: Weidenfeld & Nicolson.

Baumol, William J. (1994), 'Multivariate growth patterns: contagion and common forces as possible sources of convergence', in William J. Baumol, Richard R. Nelson and Edward N. Wolff (eds), *Convergence of Productivity*, Oxford: Oxford University Press, pp. 62–85.

Becker, Gary S. (1998), *Accounting for Tastes*, Cambridge, MA: Harvard University Press.

Berger, Brigitte (1991), 'The culture of modern entrepreneurship', in Brigitte Berger (ed.), *The Culture of Entrepreneurship*, San Francisco, CA: ICS Press, pp. 13–32.

Berger, Peter and Hsing-Huang M. Hsiao (eds) (1988), *In Search of an East Asian Development Model*, New Brunswick, NJ: Transaction Books.

Coase, Ronald H. (1988), *The Firm, the Market and the Law*, Chicago, IL: University of Chicago Press.

Coase, Ronald H. (1991), 'The institutional structure of production', Alfred Nobel Memorial Prize Lecture in Economic Sciences, reprinted in Ronald H. Coase (1994), *Essays on Economics and Economists*, Chicago, IL: University of Chicago Press, pp. 3–14.

Cyert, Richard and James G. March (1963), *A Behavioral Theory of the Firm*, Englewood Cliffs, NJ: Prentice-Hall.

Dawson, J.W. (1998), 'Institutions, investment and economic growth: new cross country evidence', *Economic Inquiry*, **36**, 603–19.

Demsetz, H. (1964), 'The exchange and enforcement of property rights', *Journal of Law and Economics*, **7**, 11–26.

Demsetz, H. (1967), 'Toward a theory of property rights', *American Economic Review*, **57**, 347–59.

Easterly, W. and R. Levine (1997), 'Africa's growth tragedy', *Quarterly Journal of Economics*, **112**, 1201–50.

Edgerton, Robert B. (1992), *Sick Societies*, New York: Free Press.

Eggertsson, Thráinn. E. (1991), *Economic Behavior and Institutions*, Cambridge: Cambridge University Press.

Greif, A. (1989), 'Reputation and coalitions in medieval trade: Maghribi traders', *Journal of Economic History*, **59**, 857–82.

Greif, A. (1994), 'Cultural beliefs and the organization of society', *Journal of Political Economy*, **102**, 912–50.

Grondona, Mariano (2000), 'A cultural typology of economic development', in Lawrence E. Harrison and Samuel P. Huntington (eds), *Culture Matters*, New York: Basic Books, pp. 44–55.

Grossman, S. and O. Hart (1986), 'The costs and benefits of ownership: a theory of vertical and lateral integration', *Journal of Political Economy*, **98**, 691–719.

Grubel, H.G. (1998), 'Economic freedom and human welfare: some empirical findings', *Cato Journal*, **18**, 207–304.

Hanson, J.R. II (1999), 'Culture shock and direct investment in poor countries', *Journal of Economic History*, **59**, 1–16.
Harrison, Lawrence E. (2000), 'Why culture matters', in Lawrence E. Harrison and Samuel P. Huntington (eds), *Culture Matters*, New York: Basic Books, pp. xvii–xxxiv.
Hirshleifer, Jack (1987), *Economic Behavior in Adversity*, Chicago, IL: University of Chicago Press.
Hofstede, Geert (1982), *Culture's Consequences*, Newbury Park, CA: Sage.
Hofstede, Geert (1991), *Cultures and Organizations*, London: McGraw-Hill.
Huntington, Samuel P. (2000), 'Foreword', in Lawrence E. Harrison and Samuel P. Huntington (eds), *Culture Matters*, New York: Basic Books, pp. xii–xvi.
Keefer, Ph. and S. Knack (1997), 'Why don't poor countries catch-up? A cross-national test of an institutional explanation', *Economic Inquiry*, **35**, 590–602.
Knack, S. (1996), 'Institutions and the convergence hypothesis: the cross-national evidence', *Public Choice*, **87**, 207–28.
Knack, S. and Ph. Keefer (1995), 'Institutions and economic performance: cross-country tests using alternative institutional measures', *Economics & Politics*, **7**, 207–27.
Knack, S. and Ph. Keefer (1997), 'Does social capital have an economic payoff? A cross-country investigation', *Quarterly Journal of Economics*, **112**, 1251–87.
La Porta, R., F. Lopez-de-Silanes, A. Shleifer and R.W. Vishny (1997), 'Trust in large organizations', *American Economic Review*, **87**, 333–8.
La Porta, R., F. Lopez-de-Silanes, A. Shleifer and R. Vishny (1999), 'The quality of government', *Journal of Law, Economics, and Organization*, **15**, 222–79.
Lal, Deepak (1998), *Unintended Consequences*, Cambridge, MA: MIT Press.
Landa, Janet T. (1994), *Trust, Ethnicity and Identity*, Ann Arbor, MI: University of Michigan Press.
Lavoie, Don and Emily Chamlee-Wright (2000), *Culture and Enterprise*, London and New York: Routledge.
Lazear, E.P. (1999), 'Culture and language', *Journal of Political Economy*, **107**, S95–S126.
Libecap, Gary (1990), *Contracting for Property Rights*, Cambridge: Cambridge University Press.
Macfarlane, Alan (1987), *The Cultural of Capitalism*, Oxford: Basil Blackwell.
Mahoney, P.G. (2001), 'The common law and economic growth: Hayek might be right', *Journal of Legal Studies*, **30**, 503–25.
Marshall, Alfred (1920), *Principles of Economics*, 8th edn, London and Basingstoke: Macmillan.
Muller, Siegfried. H. (1964), *The World's Living Languages*, New York: Ungar.
North, Douglass C. (1981), *Structure and Change in Economic History*, New York: W.W. Norton.
North, Douglass C. (1990), *Institutions, Institutional Change and Economic Performance*, Cambridge: Cambridge University Press.
North, D.C. (1994), ' Economic performance through time', *American Economic Review*, **84**, 359–68.
Norton, S.W. (1998) 'Poverty, property rights, and human well-being: a cross national study', *Cato Journal*, **18**, 233–45.
Norton, S.W. (2000), 'The cost of diversity: endogenous property rights and growth', *Constitutional Political Economy*, **11**, 319–37.
Norton, S.W. (2003), 'Economic institutions and human well-being', *Eastern Economic Journal*, **29**, 23–40.
Ostrom, Elinor, Roy Gardner and James Walker (1994), *Rules, Games and Common Pool Resources*, Ann Arbor, MI: University of Michigan Press.
Political Risk Services (2000), *International Country Risk Guide*, East Syracuse, NY: PRS Group, Inc.
Romer, P. (1986), 'Increasing returns and long-run growth', *Journal of Political Economy*, **94**, 1002–37.
Rosenberg, Nathan and L.E. Birdzell, Jr. (1986), *How the West Grew Rich*, New York: Basic Books.

Ruttan, V.T. (1991), 'What happened to political development?', *Economic Development and Cultural Change*, **39**, 265–92.
Scully, G.W. (1988), 'The institutional framework and economic development', *Journal of Political Economy*, **96**, 652–62.
Scully, Gerald W. (1992), *Constitutional Environments and Economic Growth*, Princeton, NJ: Princeton University Press.
Scully, G.W. (1997), 'Rule and policy spaces and economic progress: lessons for Third World countries', *Public Choice*, **90**, 311–12.
Temin, P. (1997), 'Is it Kosher to talk about culture?', *Journal of Economic History*, **57**, 267–87.
Todaro, Michael P. (1997), *Economic Development*, 6th edn, Reading, MA: Addison-Wesley.
Tornell, A. and Ph. R. Lane (1999), 'The voracity effect', *American Economic Review*, **89**, 22–46.

5 The effect of transaction costs in the definition and exchange of property rights: two cases from the American experience

Gary D. Libecap

Introduction: transaction costs, institutional change and economic welfare

Recent research examining cross-country differences in economic growth points to the importance of the institutional structure of a society in explaining observed variations in performance. In particular, institutions that include clearly defined and judiciously enforced private property rights appear to play key roles in promoting economic growth and welfare (North 1990; Barro 1991, 1996, 1997; Shleifer and Vishny 1993; Alston et al. 1996; Acemoglu et al. 2001a, 2001b). Given the observed importance of the property rights structure, one might predict that individuals would mobilize to bring about institutional change whenever there were net benefits of doing so. These new arrangements would allow resources to flow more easily and quickly to higher-valued uses and support increased levels of trade and investment. The resulting higher levels of economic growth would motivate the parties to more precisely define and enforce property rights.

Indeed, Demsetz (1967) and Davis and North (1972) optimistically hypothesized just such a beneficial process of institutional change. They suggested that shifts in factor and product prices and the development of new technologies would encourage individuals to refine property rights so as to take advantage of new market opportunities. Neither Demsetz nor Davis and North, however, detailed this process of institutional change, and unfortunately, as is often the case, the devil is in the details.

Although Davis and North described some of the constituencies and organizations that would be involved, the underlying bargaining and any problems it might encounter were not specifically addressed. Subsequent research has shown that the timing and outcome of institutional change are critically affected by the transaction costs encountered. Williamson's (1979) work outlined the importance of considering transaction costs in explaining observed institutions and understanding economic behavior. Additionally, North's (1990) and Barzel's (1989) subsequent insights into transaction costs, property rights and economic growth led him to take a more measured view of the prospects for institutional change. He noted that economic history was characterized by the persistence of

seemingly inefficient property rights structures, a slow and halting process of institutional change, and limited economic growth. Libecap (1986, 1989) developed the implications of transaction costs for property rights definition to resolve common-pool problems during economic development, and Libecap and Smith (2003) used the Demsetz hypothesis of institutional change to analyse the development of oil and gas property rights and regulation in the United States from the mid-nineteenth century to the present. They found that although there was a general pattern of progressive property rights change in response to new market conditions, organizational arrangements and regulatory policies that included many apparent inefficiencies were adopted. These institutions could only be explained through consideration of the transaction costs of negotiating new rights structures.

Once those transaction costs were considered, Libecap and Smith argued that it was no longer possible to draw efficiency judgments about observed conditions. As Demsetz (1969) noted, it was not appropriate to compare the efficiency of existing conditions with an alternative, hypothetical example that could only emerge in a world of zero transaction costs. Transaction costs were real factors that influenced the development of actual institutional arrangements, and as such, had to be accounted for in any conclusions about the efficiency of observed arrangements.

In sum, transaction costs mold bargaining for changes in property rights, influencing the positions of the parties involved and the nature of the institutions that ultimately result. Accordingly, understanding the process of institutional change, which seems so critical in explaining differences in economic growth, requires attention to transaction costs – their nature and sources and why they may be so intractable. In this chapter, I briefly summarize some issues regarding transaction costs and then turn to two empirical examples in the United States where transaction costs have importantly affected the transfer of property rights to higher-valued users. One example is the persistence of seemingly inefficient small farms on the American Great Plains. Another example is the continued allocation of water rights to agriculture in the semi-arid American West, despite efforts to reassign water to urban and environmental uses where it has much greater value. Examination of these two cases helps clarify the nature of transaction costs and how they have slowed resource allocation, even in the presence of substantial changes in relative prices and technology.

Transaction costs and changing property rights

Transaction costs are essentially information costs that include the search, negotiation and enforcement costs in both private efforts to define and enforce property rights to resources and in government efforts to devise and implement ownership policies. Among other factors, transaction costs are a

function of the nature and distribution of information about the resource, its physical characteristics and value, and the size and heterogeneity of the negotiating group.

A lack of knowledge about the resource increases uncertainty about the magnitude and distribution of the benefits and costs of developing and refining property rights. This situation makes it more difficult for the negotiating parties to agree to changes in the property rights structure. They cannot know with much accuracy how the proposed change in ownership will affect resource value and use or how each party will benefit from (or be harmed by) the new arrangement. More contingencies must be considered, more options weighed, and potentially, more disagreement encountered over suggested alternatives. These conditions raise the transaction costs of negotiating and implementing institutional change. Similarly, if some parties have better information about the effects of the proposed arrangement, but cannot convey it credibly to the other parties, then there will be conflicting views about the net effects of property rights change and how the benefits and costs are likely to be distributed. The parties will only agree to voluntary changes in the property rights structure if they anticipate improvements in their welfare, relative to the status quo. If such individual assessments are difficult to determine, then bargaining disputes will slow or block agreement on a new ownership arrangement until information is more broadly distributed and accepted. Such information problems and their effects on property rights are examined in Libecap and Wiggins (1985) and Wiggins and Libecap (1985) with respect to addressing production externalities in American oil and gas reservoirs. These authors found that new arrangements to address serious common-pool externalities often took seven years or more to be approved, while in the meantime, there were serious rental losses from excessively rapid exploitation.

Of more concern here are issues of compensation when there is limited information about the resource and the results of a new property rights arrangement. Under these circumstances, it will be difficult to assess the claims for compensation made by those who expect to be harmed by modification of property rights. Compensation requires agreement on those who have legitimate claims and on the value of the 'harms' imposed on them. Some losses may be very difficult to compensate. For example, politicians whose political base is dependent upon a particular allocation of property rights, wealth and associated political influence, might find little attraction in a reallocation that weakens or eliminates their political support. Whether or not this loss can be valued, whether the other parties will deem it legitimate for compensation, and the form of compensation that might be paid can be formidable issues to be addressed.

In the empirical case of political resistance to small-farm consolidation in the Great Plains, politicians and administrators whose futures would have

been limited by sharp rural population declines had incentives to intervene to slow the process and to protect their positions. Politicians took these actions even though there were aggregate efficiency losses from doing so and income transfers from other taxpayers were necessary to support small farmers. Similarly, opposition to potentially valuable water transfers in the semi-arid West illustrates how transaction costs can be increased when third parties anticipate that they will be negatively affected by a redistribution of property rights and not compensated. Proposed reallocation of water from agriculture to urban uses brings concern about uncompensated community effects if agriculture subsequently declines. Indeed, these considerations underlie much of the resistance to water transfers to higher-valued uses outside of agriculture. For many third parties, local property owners, agricultural support industries, and rural politicians and administrators, there is no precedent or obvious means of compensation for their losses if the agricultural base is eroded. These parties may be in a position to delay, modify or block institutional change that otherwise has broad aggregate benefits.

Large, mobile, unobservable resources also are associated with greater costs of measuring and enforcing property rights. Enforcement is essential for ensuring the integrity of any property system. In the absence of effective enforcement, property rights are not exclusive. Without exclusivity property 'owners' will be neither the sole decision makers over resource use nor will they be able to capture the net benefits of their decisions. These conditions reduce time horizons for exploitation and trade and diminish incentives for investment. For example, private ownership of migratory ocean fish is virtually impossible because of the high cost of defining and securing property rights. As a result, the range of institutional options for investing in fish stocks and protecting them from overharvest is quite limited.

Also, enforcement costs can be raised if property rights are not clearly defined, as is the case of fisheries, or are overlapping, as in the water case described in this chapter. Under either of these circumstances there likely will be conflicting 'owners', and sorting out who has the right to the resource and who will bear the costs and benefits of institutional change will be complicated and costly. Because overuse is common in such cases, the more precise definition and enforcement of property rights means that some parties will be excluded. As a result, institutional change may make some parties worse off, and they may not have grounds for compensation. These individuals, naturally, have incentives to oppose property rights change. For example, in the case of water, there are often multiple claimants since it is re-usable. If those with a senior or priority property right agree to a more precise definition of rights, which allows for the transfer of water to users outside the region, however, there may be less 'return flow' and accordingly, less water for junior rights holders. Although, in theory these additional parties can be

compensated for their losses, their standing may be disputed and the value of their damages rejected. If this is the case, they have incentives to resist property rights change. Even if their concerns are recognized as legitimate, there must be agreement on the nature and amount of compensation to be granted and the identities of the parties who will pay. These issues involve information search and discussion, and are apt to be bound with disagreement – all of which raise transaction costs.

Additionally, higher-valued resources will have more competing parties, which means more claims and concerns that must be addressed in negotiations, increasing transaction costs.[1] The greater the heterogeneity of the parties in terms of objectives, production costs, access to information, and other factors, the more difficult it is to reach consensus on actions to take and on the distribution of costs and benefits. With heterogeneous parties the individual net benefits of assigning rights vary. In private and political negotiations to change property rights, each of the bargaining parties attempts to maximize individual net returns by increasing its share of the aggregate gains and reducing its portion of the costs. These actions, however, can limit and delay the property rights that are adopted. These issues are clearly raised in the opposition of third parties to the reallocation of property rights to land in the Great Plains and to water in the semi-arid American West.

Negotiations to change property rights involve offers and counter offers, and in the political arena, these negotiations involve logrolling. As noted above, side payments in the form of preferential property rights, regulation or monetary transfers can be proposed as part of logrolling trades to mitigate opposition from those who otherwise expect to be harmed. Reaching agreement on the identities of the affected parties, the magnitudes of the impacts, the amount of compensation and the parties to pay are complicated issues that slow the process of institutional change.

In the presence of transaction costs, Coase's (1960) analysis implies two measures to promote changes in property rights. One action is to determine whether there is a feasible way to decrease the costs of transacting among agents in order to facilitate private agreement. The other option is to devise government policies, including taxes and regulatory actions, to promote institutional change. Both measures are costly and are affected by the existing distribution of property rights. Current rights holders, whether informal or formal, and related third parties, whose welfare depends on the existing allocation of ownership and resources, are critical constituencies in any institutional change. These individuals have a vested interest in the status quo and will oppose change unless they can be made better off under the proposed arrangement.

It may not be possible, however, to improve their welfare and still maintain the advantages of the institutional change, especially if there is disagreement

over compensation. Groups with vested interests may have advantages in political bargaining relative to others through lower costs of collective action. Their current position in the system binds them together to make them a relatively cohesive bargaining group. They also may have beneficial ties to established political processes and leaders. These advantages make vested interests effective negotiators and lobbyists, biasing institutional change toward protection of the status quo.

Transaction costs and political opposition to changes in property rights in the American Great Plains

US land distribution policy in the late nineteenth and early twentieth centuries resulted in farms that were too small to be economically viable over the long term in the American Great Plains. Most federal land laws were enacted prior to major settlement of the region when there was little understanding of its semi-arid climate or of appropriate agricultural techniques and farm sizes for dry areas. The Homestead Act was the major federal land law which allowed for private claiming and patenting of up to 160–acre (later, 320-acre) plots of federal land. Under the Homestead Act, there was a proliferation of small farms throughout the Great Plains: 1 078 123 original homestead entries were filed covering 202 298 425 acres in western Kansas, Nebraska, the Dakotas, eastern Colorado and Montana between 1880 and 1925.[2] Homestead property rights allocations worked well in northern agriculture, east of the 98th meridian. There were no important economies of scale in grain production, and there was sufficient rainfall (usually above 30 inches a year), high soil quality, and familiar conditions, allowing farmers to use knowledge gained in the East or in Europe. As migrants moved across the frontier, they transplanted farming practices, crops and farm sizes appropriate in their places of origin, but not, as it turns out, for the Great Plains.

West of the 98th meridian, the beginning of the North American Great Plains, things were different. It was dry. Precipitation varied across the region, and was often below 15 inches a year. More critically, precipitation was highly variable, resulting in periodic severe droughts, which slashed crop yields and farm incomes and placed small homesteads at risk. Due to their small size, homesteads had little cushion to compensate for declining output and were unable to diversify into crops and livestock less affected by drought (Hansen and Libecap 2004a). Nevertheless, at the time of settlement of the Great Plains, there was little scientific understanding of agriculture in semi-arid lands, and importantly, except for a drought in the northern plains from 1917 to 1921, most migration occurred during a period of abnormally high rainfall.

As early as 1879, John Wesley Powell in his *Report on the Arid Lands of North America* recommended farm sizes of 2560 acres for the region, but

Great Plains politicians were unanimous in their opposition, and there were no major adjustments in federal land laws beyond the 1909 change that allowed for 320-acre homesteads.[3]

The body of scientific knowledge regarding the region's climate and appropriate agricultural techniques and farm sizes developed only slowly (Libecap and Hansen 2002). Even as new information emerged that suggested larger allocations might be in order, the science remained unclear and there were powerful political pressures for maintaining the existing piecemeal division of land. So long as precipitation levels remained relatively high, small homesteads were sufficiently productive to earn sufficient income to sustain a family and support the notion of small-farm settlement.[4] There also were no long-term records of precipitation or accepted predictive models of drought occurrence to undermine the sense of well-being that permeated the Great Plains. Indeed, during early settlement periods, there was a belief (the 'rain-follows-the-plow' doctrine) that the climate had been permanently changed by cultivation. When this assessment was later discredited by the reappearance of regional droughts, new cultivation techniques (the so-called 'dryfarming doctrine') were developed and asserted to have 'defeated' drought through water storage (ibid.).

Accordingly, it is understandable why local and regional politicians were skeptical about claims that farms needed to be larger on the Great Plains. They sought to maintain the dense settlement patterns and population levels that had supported economic development and provided opportunities for advancement from local and state to federal political office. In particular, the number of members in the House of Representatives was based on population, and state politicians were very reluctant to allow populations to decline, reducing the positions available to them in the Congress. These concerns were mirrored at the local level, since representation of rural counties in state legislatures was also based on population. Additionally, county tax revenues and local commercial activities depended upon vibrant rural communities. As we shall see, these political interests subsequently raised the transaction costs of farm consolidation.

The long drought of the 1930s and related severe wind erosion (the 'dust bowl'), however, demonstrated the vulnerability of small homestead farms. Wheat yields collapsed along with farm incomes and the viability of small homesteads. With the effects of the Great Depression and drought in the Great Plains, the federal government initially attempted to encourage outmigration and the formation of larger farms on the Great Plains through the Resettlement Administration. Although farm consolidation was starting through private purchase and through bank foreclosure of mortgages and subsequent sale, the process was slow, and many small farmers resisted sale. Federal officials concerned about rural poverty suggested that the Great Plains could

sustain only two-thirds of its 1930 population (see Thornthwaite 1936).[5] Writers in other government agency reports argued that only much larger farms could be economically sustainable in the region (see Hansen and Libecap 2004b). Accordingly, using the superior resources of the federal government, the Resettlement Administration began to purchase small farms to speed the process of consolidation. Nevertheless, the policy had limited success. In most cases the government's purchase and resettlement programs were sharply resisted (Worster 1979, pp. 42–6). Eventually they were abandoned and replaced by a new system of relief payments and subsidies to maintain small farms. The Resettlement Administration was renamed and restructured as the Farm Security Administration.

Great Plains politicians feared a loss in farm population, the related deterioration in local economic activity, and decline in national political influence should rural populations plummet with out-migration. As noted above, the number of representatives in the House was at stake, as were property values in rural communities and related investment in schools and other infrastructure. It is not obvious how local and regional politicians could have been compensated for resettlement policies that would have eliminated their offices. Also, small farmers were reluctant to sell. They only had to cover the opportunity costs for variable inputs, labor and capital, and their human capital was linked to agriculture with few other options in the region other than migration. Even migration was not attractive during the Depression when unemployment levels were high in urban areas. Hence, farmers sought to stay on their farms as long as possible. Local community leaders joined them in lobbying for subsidies and regional politicians were only too happy to respond.

Small farmers, or 'family farmers', became a critical constituency in developing policies that raised the transaction costs of farm consolidation. They had no obvious alternative employment in agriculture in the 1930s or later, and apparently there was no broad political support elsewhere in the country for sufficiently high payments to induce small farmers to leave agriculture and retire. There was no feasible compensation for regional politicians and related bureaucratic officials whose positions would have been threatened by mass out-migration.

Accordingly, the real thrust of government policy in the 1930s in the Great Plains was to sustain family farms via subsidies. Neither local politicians nor officials of the Department of Agriculture wanted to see a dramatic loss of farmers in the region. In its 1938 *Yearbook of Agriculture, Soils and Men*, the agency noted the debate over whether to move farmers out of farming or to subsidize them, and sided with the latter: 'it is wise to keep a large rural population'(pp. 3–4). The department stood to lose much of its constituency in the region. Clawson et al. (1940, pp. 42–8) claimed that eliminating farms

of less than 300 acres in eastern Montana would reduce the number of farms by 76 per cent. But they doubted that many would be willing to accept such drastic steps. They still called for the elimination of 50 per cent of the farms in the region from 1928–35 levels, and predicted it would take 30 years to do so with considerable government assistance.

As a result, an elaborate system of subsidies was set up through the Farm Security Administration, the Works Progress Administration, the Farm Credit Administration and the Federal Emergency Recovery Administration (FERA) to maintain farms and rural populations. The major historian of the Dust Bowl, Donald Worster (1979, pp. 131–5) estimated that three out of four farmers in the region received federal aid. Johnson (1947, p. 190) noted that in some areas as many as 80 per cent of the farmers were on relief. A March 1935 survey indicated that up to 40 per cent of farm families in the Texas panhandle, over 50 per cent in southeastern Colorado, and between 33 and 50 per cent in southwest Kansas were dependent on government payments. Between September 1933 and August 1935, FERA granted $32 666 370 to Colorado, Kansas and Oklahoma for relief. Those who were not on relief were able to stay on their farms mainly because of crop adjustment payments from the Agricultural Adjustment Administration (AAA). Between 1933 and 1936, total federal aid averaged $223/person in 72 southern plains counties.[6]

The subsidies, however, effectively raised the transaction costs of purchasing and consolidating small farms, helping to sustain otherwise non-viable units, delaying the adjustment toward larger and more efficient farm sizes.[7] Libecap and Hansen (2004a) argue that this property rights distribution encouraged serious environmental damage through wind erosion since small farms were less likely to invest in erosion control. Most of the gains of doing so were captured by downwind farmers and the lack of erosion control contributed to the dust bowl. Other efficiency losses likely included too-intensive land use since small farms could not afford to leave land in fallow. Government subsidies made it possible for small farmers to stay on the land, delaying the reallocation of property rights to more efficient units.

Gradually, farms were consolidated, especially as farmers died or retired and their heirs preferred not to maintain the farm. The process took approximately 60 years (Hansen and Libecap 2004a). Without these subsidies, more small farmers would have migrated from the Great Plains in the 1930s, farm consolidation would have occurred more rapidly, and more efficient land-use practices would have been put into place. Further, the infrastructure investments, such as schools, county governments and roads, necessary to maintain an unsustainably large population could have been reduced.[8]

Table 5.1 describes the nature of the farm-size adjustment process on the Great Plains. It provides census data for two Great Plains states, Colorado and Montana for 1920 and 1982. In 1920, mean farm size in the two states

Table 5.1 *Farm-size adjustment on the Great Plains, 1920–1982: Montana and Colorado*

	Montana	Colorado
1920		
No. of farms		
Less than 100 acres	4 350	15 294
100–499 acres	35 723	33 750
500–999 acres	11 982	7 482
Over 1000 acres	5 622	3 408
Total	57 677	59 934
Mean farm size (acres)	608	408
Std deviation	1.402	1.119
CV	2.30	2.74
1982		
No. of farms		
Less than 100 acres	5 593	9 252
100–499 acres	4 808	7 761
500–999 acres	2 640	3 337
Over 1000 acres	10 529	6 761
Total	23 570	27 111
Mean farm size (acres)	2 568	1 237
Std deviation	2.359	2.071
CV	0.92	1.67

Source: US Census.

was 408 and 608 acres, respectively. Most of the farms were less than 500 acres, and there was considerable heterogeneity in farm sizes as indicated by the coefficient of variation (CV), which was 2.7 for Colorado and 2.3 for Montana. By 1982, however, mean farm size was much larger at 1237 and 2568 acres. Further, the variance in farm size had declined. The CV was 1.67 for Colorado and 0.92 for Montana. Farm sizes finally had coalesced around the mean.

Transaction costs and the reallocation of water rights from agriculture to urban and environmental uses

In semi-arid western US states, most water historically was used in agriculture. This remains true even today, as irrigation withdrawals account for approximately 70–80 per cent of annual water use in the West. In the past 30

years, rapid population growth, urbanization, a rise in the contribution of manufacturing and services along with a relative decline in agriculture, and increased environmental concerns have led to increased demand for water in urban areas and associated pressure to reallocate it from agriculture. Even so, permanent transfers from agriculture have been limited in many western states. A principal issue has been concern about the effects on agricultural economies and communities from water transfers. Opposition from rural interests has sometimes led to restrictions on transfers. Throughout the West, the reallocation of water from rural to urban areas has been controversial, and these controversies have raised the transaction costs of redistributing water.

Although some transactions have occurred through market trades, most reallocations have involved exchanges between institutional entities in response to exogenous mandates, legislation or court rulings. The transaction costs of reallocating water are very high and vary across the states, leading to sharp differences in water prices between agricultural and urban uses. For instance, groundwater for farming near Marana, Pima County, Arizona costs approximately $25 per acre-foot (approximately 325 000 gallons), whereas the same water for urban use costs $200. In addition, there are few opportunities for arbitrage. Similarly, in recent efforts to secure Imperial Irrigation District water, San Diego offered $225 per acre-foot for water that farmers used for $15.50. Even more dramatically, while farmers in the Imperial Irrigation District paid $13.50 per acre-foot in 2001, a development near the South Rim of Grand Canyon National Park was prepared to spend $20 000 per acre-foot to deliver the same Colorado River water (Glennon 2002).

The economic theory of institutional change as outlined above suggests that with such opportunities for trade, water law and related legal institutions will respond to lower the transaction costs of transferring water from agriculture. Yet as noted, recent work on institutional change reveals that the process is complex and can be derailed by information problems, distributional concerns, entrenched political constituencies, and third-party effects that cannot be completely compensated. According to Coase (1960), optimal resource allocation occurs when there is: (i) costless negotiation, (ii) fully-defined property rights and (iii) the absence of wealth effects. Each of these conditions is violated in some manner with water. Transaction costs in negotiating water transfers are high because there are multiple parties involved with heterogeneous interests; they have incomplete or overlapping property rights with important third-party effects; their positions on transfers are affected by information problems; and there may be distributional consequences from transfers that are difficult to compensate. These problems are examined below. Similarly, Posner (1977) details three criteria for an efficient system of property rights: complete definition, exclusivity and transferability. Both Demsetz (1967) and Posner (1987) hypothesize that these criteria are more

likely to be met as resource values rise, but they do not appear to exist for many water rights despite rising water values. The work of North (1990) and Williamson (1979) suggests that a variety of issues can impede institutional change in water.

Externalities from water transfers that raise transaction costs

Whether selling long-term water rights or leasing in the short term, a third party besides the buyer and seller is nearly always affected. These third-party effects are the most important factors in raising the transaction costs of water transfers (Young 1986). One effect arises from issues of return flow. When water is diverted from a stream or pumped from an aquifer for local use, much of it returns for adjacent use through return flow. When water is sold over long distances, however, the return flow is disrupted, affecting other users of that water source. Moreover, when water is diverted from a stream, the remaining water often has higher salinity, which reduces its value to farmers and other users. Further, the recharge or level in an aquifer is affected when water is pumped for outside uses. Should that level fall below a certain point, permanent harm could be inflicted on the aquifer, affecting all users.

If transaction costs were very small, third parties could be compensated for their harm. But measuring harm accurately and identifying legitimate claims can be very difficult. These third-party effects have limited the development of water markets. Water markets have formed where such effects are low, either due to a limited number of potential third parties or to a unique geography that makes return flow easy to quantify, track and measure. Water markets have been more successful when pumping from groundwater sources for this reason. The number of parties with land above an aquifer is finite and easily determined. This simplifies the quantification of third-party effects. Water markets have been much less successful using surface water from a river since the number of down-river parties is nearly infinite. These issues are compounded when the river passes through more than one state, country or Indian reservation – which all have sovereign powers.

In order to address externalities, many states have enacted 'public harm' provisions that stipulate that water may not be transferred if doing so harms the public good. These statutes can add to uncertainty and transaction costs because they seldom stipulate the public interest, the damage to it, or identify the responsible parties. For example, California's public harm statute puts the burden of proof on an applicant for a water transfer to prove that no harm will be done. This of course raises the cost of transfer. Even states with no such stipulation, such as New Mexico, have not laid out who is responsible or who may make such claims (Bokum 1996; Klein-Robbenhaar 1996). The imprecision in the statutes raises uncertainty as to the direction of the law and hence, transaction costs.

While externalities associated with return flow and aquifer recharge are difficult to measure, at least in theory, compensation is possible. Other third-party effects are even more difficult to compensate. One is equity. Increases in water values and the development of water markets have led to the accrual of substantial rents to sellers, upsetting status quo wealth distributions. While voluntary market transactions require that all parties be made better off, equity questions have been raised regarding the appropriateness of these windfalls. So long as urban users question the equity of these rents, current agricultural water rights holders may be reluctant to enter into or form water markets out of fear of future appropriation or taxes by the state. Indeed, political reaction to wealth windfalls from naturally occurring resources, such as water, has resulted in legislation limiting the gains possible from trade in some states. Other equity concerns arise when water transfers lead to substantial increases in wealth outside of the area of water origin. For example, if water is shifted from rural to urban areas to facilitate economic growth, rural areas and economies may languish whereas urban economies may benefit.

As described earlier, compensation requires identifying the relevant parties and measuring the negative impacts. While the landowners may be compensated through the sale of water rights, other parties, such as farm supply and support businesses, or farm labor may suffer losses (Howe et al. 1990). Further, county budgets may be adversely affected as tax bases shrink when municipalities purchase farmland for water and retire the farms from agriculture. If the property purchased lies outside the county in which the city resides, the adverse fiscal distributional effects are even greater, with the rural county losing its tax base and the urban county gaining. While one can imagine this situation being adjusted by state legislatures with revenue sharing, opposition may be strong from urban communities, especially if the costs of securing the land and water had been high. Finally, rural political influence may be lost when large water transfers are made, and there is no clear mechanism for compensation, especially to local politicians (Nunn 1985). This problem is similar to that which occurred regarding small-farm consolidation in the Great Plains. For all of these reasons, some parties may have an interest in securing legal restrictions on the transfer of water from agriculture even though they may be aggregately socially beneficial.

Concerns from rural areas regarding water transfers have resulted in political action to block or limit water transfers from agriculture. For example, in Arizona the cities of Prescott, Scottsdale and Tucson purchased farm land between 1960 and 1985 to secure additional water by retiring the acreage and reallocating the water to urban use. In 1990, the state legislature at the behest of rural interests banned further transportation of water from rural to municipal areas.

Institutional arrangements that raise transaction costs of water transfers

A second factor affecting transaction costs is the institutional structure. Most agricultural water is administered by irrigation districts or Indian tribes. Few irrigation districts have supported water transfers to urban areas, and some, such as California's Imperial Irrigation District, which uses most of the state's allocation of Colorado River water, have steadfastly resisted reallocations.

Thompson (1993) has identified nine possible reasons for this reluctance to transfer water, all of which increase the transaction costs of water exchange. His list includes the following:

1. problems with incentive structures within districts, since water ownership and decision-making authority varies across districts and often involves regulation by district governance boards, so that motivation for transfers differs across districts;
2. possible negative externalities to non-sellers as water is shipped elsewhere;
3. possible harm to the selling communities if the agricultural economy is diminished;
4. conflict with local ethics from sale of a common resource;
5. possible increases in internal water prices from water sales;
6. increased administrative costs in evaluating and monitoring the effects of sales and arbitrating disputes;
7. increased uncertainty of water supplies, especially to junior rights owners during drought;
8. loss of managerial authority with irrigation districts if agriculture diminishes; and
9. interest in stability, which could be disrupted if substantial amounts of water are sold.

Each of these factors affects the incentives of irrigation districts to release water to urban areas.

Even beyond incentives within a single district, the existence of multiple irrigation districts with different regulations and procedures raises transaction costs compared to situations where a single district with a single set of rules covers a large region. These points are raised in Carey and Sunding's (2001) comparison of the California Central Valley Project with multiple districts and the Colorado Big Thompson Project with a single district. Further, if water is transferred from Indian reservations, federal law applies and there are different governance structures across the various tribes.

A third institutional factor affecting transaction costs is the existing drought-rationing mechanism. In all western states, the prior appropriation system

governs rights to surface water. The senior, or earliest, diverter has the first claim. In times of drought, junior appropriators may be cut off. This system complicates water exchanges because the certainty of obtaining water varies according to the priority of its origin. Water sold by junior appropriators may be withheld by senior appropriators during drought periods, blocking or at least limiting the transfer of water to urban users.

Another institutional factor affecting transaction costs is return flow rights. In the Colorado Big Thompson project all return flow rights are held by the water district, which effectively eliminates all downstream third-party effects, thus lowering transaction costs and uncertainty for exchange. By contrast, where return flow rights are held by other districts or water users, they will be affected by water transfers by upstream owners. Depending on state law, upstream owners may have no obligation to consider the impact of a transfer on downstream owners. Water courts for arbitrating disputes could conceivably address the issue and suggest compensation, but return flow effects are also affected by precipitation levels and perhaps, multiple upstream and adjacent groundwater users. Accordingly, quantifying the harm caused on downstream owners can be daunting.

Uncertainty and information problems that raise transaction costs of water transfers

A third source of transaction cost in water transfers is uncertainty and related information problems. In the semi-arid West, both agriculture and urban development require access to water. Drought is a major source of uncertainty because there are no effective predictive models. Accordingly, transfers of some agricultural water to urban areas could place that agriculture at risk if a serious drought occurs. This condition makes it difficult for farmers to release part of their water rights since they lose the cushion that could protect them during drought. There are other information problems associated with valuing water. Beneficial use requirements make leasing of water rights or short-term transfers risky because these actions could be interpreted as evidence of a lack of beneficial use and lead to a loss of water rights. States have been reluctant to allow interstate transfers when future demands are uncertain. Such transfers could cap future economic growth. This same concern affects rural communities. There is also uncertainty over the scope of water rights. Many states have not adequately defined rights in a way that reduces uncertainty over who controls the water and for what purposes it can be used.

Despite the rise in water values, there is no clear trend in institutional development to facilitate transfers from agriculture. DuMars and Tarlock (1989a, 1989b) find adaptive changes in water law historically due to judicial rulings, legislative statutes, administrative actions, and federal intervention via large interstate water projects. Anderson's (1985) case study in Hawaii,

however, documents slow progress in water market development. Further, Tarlock (1991) observes legal changes in western states that have made water transfers more difficult. He argues that the *ad hoc* patchwork of legal rulings and laws have made water transfers more complex. Nunn (1985) argues that institutional change is slowed by strong agricultural lobbies. Getches (2001) describes gradual changes in water law in response to new urban and environmental demands, but he notes that many states have enacted legislation limiting or restricting water transfers.

Concluding remarks: transaction costs and property rights change

In 1967, Harold Demsetz advanced the thesis that development of property rights flows from underlying changes in the relative prices of goods and the technologies that are used to produce them. Institutional change takes place whenever there are net gains from doing so, and through this process new property rights are created to replace those no longer attuned to economic conditions. Demsetz noted, however, that the actual arrangement adopted depends upon transaction costs. The analysis here expands on the importance of transaction costs, which was not really developed by Demsetz: 'In general, transacting costs can be large relative to the gains because of "natural" difficulties, difficulties in trading, or they can be large because of legal reasons' (Demsetz 1967, p. 348).[9] The two empirical examples examined in this chapter, the slow adjustment in farm sizes toward more efficiently-sized units and resistance to the reallocation of water rights to much higher-valued uses, illustrate the importance of transaction costs in property rights change. Transaction costs arise from a variety of sources, but they are fundamentally due to information problems.

In these two examples, transaction costs arose from either (i) situations where the parties who must relinquish or modify their rights cannot be fully compensated for doing so or (ii) situations where third parties to the transfer are negatively affected and are not compensated. In both instances, these parties have incentives to resist property rights change. Compensation is incomplete because of the difficulty in valuing the harms imposed and because of a lack of legal standing for compensation. In the case of politicians whose jurisdictions have depended upon maintenance of rural populations, either by supporting small farms or by retaining water in agriculture, there is no ready compensation that could be developed legally to cover the loss of political office if farms were to consolidate or water were to be transferred out of agriculture. This issue similarly applies for bureaucratic officials whose constituents were in agricultural regions threatened by farm consolidation or water transfer and population loss. Diminished constituencies could mean the loss of administrative mandate and justification. Other compensation issues arose from measurement problems and the political acceptability of determining how much to pay other third

parties affected by water transfers or farm consolidation. Since these issues were not and have not been adequately resolved, farm consolidation was slowed by a deliberate policy of small-farm subsidization and water transfers from agriculture to urban and environmental uses limited by legal and institutional restrictions. These cases suggest that the information problems underlying transaction costs, especially as to how they affect compensating the various parties involved, deserve more attention if we are to understand why property rights institutions do not change in a manner that appears to support higher-valued uses and economic growth.

Notes

1. At the same time, however, more-valuable resources offer greater returns from changes in property rights to allow for higher-valued uses. Higher returns can offset the higher transaction costs involved.
2. US General Land Office, *Annual Report of the Commissioner of the General Land Office*, fiscal years 1880–1925. The calculations are for state totals.
3. See analysis of the political economy of federal land law and settlement of the Great Plains in Hansen and Libecap (2004a).
4. Libecap and Hansen (2002) calculate the returns to small homesteads under various production conditions.
5. Thornthwaite (1936, pp. 243–5) called for the slow removal of 900 000 people or 210 000 families. He presented numbers of 'surplus families' by state: North Dakota 7360, Montana 12 610, Colorado 2580, Texas 12 200, Oklahoma 2930, Kansas, 6100, Nebraska 4930 and South Dakota 4640.
6. Thornthwaite (1936, p. 246) stated: 'It is evident that many of the farmers have been able to remain on their land only through a succession of loans'. Johnson (1937, p. 162) stated that failing farms have appealed to Congress for seed and feed loans and other relief. The total of feed and seed loans and relief aid poured into some counties since 1929 exceeded the purchase value of the dryfarming land.
7. Saunderson et al. (1937, p. 18) were critical of the effects of government relief which delayed adjustment toward more viable farm units.
8. In many if not most rural Great Plains counties, population peaked by 1930 so that by the end of the century population levels were at those envisaged by the early Resettlement Administration.
9. For similar examination of the development of oil and gas property rights and regulation, see Libecap and Smith (2003).

References

Acemoglu, D., S. Johnson and J. Robinson (2001a), 'The colonial origins of comparative development: an empirical investigation', *American Economic Review*, **91** (5), December, 1369–401.

Acemoglu, D., S. Johnson and J. Robinson (2001b), 'Reversal of fortune: geography and institutions in the making of the modern world income distribution', Working Paper, Economics Department, Massachusetts Institute of Technology, Cambridge, MA.

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, T. (1985), 'The market alternative for Hawaiian water', *Natural Resources Journal*, **25** (4), 893–910.

Barro, R. (1991), 'Economic growth in a cross section of countries', *Quarterly Journal of Economics*, **106** (2), May, 407–33.

Barro, R. (1996), 'Institutions and growth: an introductory essay', *Journal of Economic Growth*, **1** (2), June, 145–8.
Barro, Robert (1997), *Determinants of Economic Growth: A Cross-Country Empirical Study*, Cambridge, MA: MIT Press.
Barzel, Yoram (1989), *Economic Analysis of Property Rights*, New York: Cambridge University Press.
Bokum, C. (1996), 'Implementing the public welfare requirement in New Mexico's water code', *Natural Resources Journal*, **36** (4), 681–713.
Carey, J.M. and D.L. Sunding (2001), 'Emerging markets in water: a comparative institutional analysis of the Central Valley and Colorado–Big Thompson projects', *Natural Resources Journal*, **41** (2), 283–477.
Clawson, M., M.H. Saunderson and N.W. Johnson (1940), 'Farm adjustments in Montana, study of Area IV: its past, present, and future', *Montana Agricultural Experiment Station Bulletin*, No. 377, Bozeman, January.
Coase, R.H. (1960), 'The problem of social cost', *Journal of Law and Economics*, 3, October, 1–44.
Davis, Lance E. and Douglass North (1972), *Institutional Change and American Economic Growth*, New York: Cambridge University Press.
Demsetz, H. (1967), 'Toward a theory of property rights', *American Economic Review*, **57**, 350–59.
Demsetz, H. (1969), 'Information and efficiency: another viewpoint', *Journal of Law and Economics*, **12** (1), 1–22.
DuMars, Ch.T. and D.A. Tarlock (1989a), 'New challenges to state water allocation sovereignty: symposium introduction', *Natural Resources Journal*, **29** (2), 331–46.
DuMars, Ch.T. and D.A. Tarlock (1989b), 'A survey of the evolution of Western water law in response to changing economic and public interest demands', *Natural Resources Journal*, **29** (2), 347–87.
Getches, D.H. (2001), 'The metamorphosis of western water policy: have federal laws and local decisions eclipsed the states' role?', *Stanford Environmental Law Journal*, **20** (1), 3–72.
Glennon, Robert (2002), *Water Follies: Groundwater Pumping and the Fate of America's Fresh Waters*, Washington, DC: Island Press.
Hansen, Z.K. and G.D. Libecap (2004a), 'The allocation of property rights to land: U.S. land policy and farm failure in the Great Plains', *Explorations in Economic History*, **41** (2), 102–29.
Hansen, Z.K. and G.D. Libecap (2004b), 'Small Farms, externalities, and The Dust Bowl of the 1930s', *Journal of Political Economy*, forthcoming.
Howe Ch.W., J.K. Lazo and K.R. Weber (1990), 'The economic impacts of agriculture to urban water transfers on the area of origin: a case study of the Arkansas River Valley in Colorado', *American Journal of Agricultural Economics*, **72** (5), 1200–1204.
Johnson, Sh.E. (1937), 'Land use readjustments in the northern Great Plains', *Journal of Land and Public Utility Economics*, **8**, 153–62.
Johnson, Vance (1947), *Heaven's Tableland: The Dust Bowl Story*, New York: Farrar, Straus & Co.
Klein-Robbenhaar, J.F. (1996), 'Balancing efficiency with equity: determining the public welfare in surface water transfers from Acequia communities', *Natural Resources Journal*, **36** (1), 37–58.
Libecap, G.D. (1986), 'Property rights in economic history: implications for research', *Explorations in Economic History*, **23**, 227–52.
Libecap, Gary D. (1989), *Contracting for Property Rights*, New York: Cambridge University Press.
Libecap, G.D. and Z.K. Hansen (2002), '"Rain follows the plow" and dryfarming doctrine: the climate information problem and homestead failure in the upper Great Plains, 1890–1925', *Journal of Economic History*, **62** (1), 86–120.
Libecap, G.D. and J.L. Smith (2003), 'The economic evolution of petroleum property rights in the United States', *Journal of Legal Studies*, forthcoming.

Libecap, G.D. and S.N. Wiggins (1985), 'The influence of private contractual failure on regulation: the case of oil field regulation', *Journal of Political Economy*, **93**, 690–714.
North, Douglass (1990), *Institutions, Institutional Change, and Economic Performance*, New York: Cambridge University Press.
Nunn, S.C. (1985), 'The political economy of institutional change: a distribution criterion for acceptance of groundwater rules', *Natural Resources Journal*, **25** (4), 867–92.
Posner, Richard A. (1977), *Economic Analysis of the Law*, 2nd edn, Boston, MA: Little, Brown.
Posner, R.A. (1987), 'The law and economic movement', *American Economic Review*, **77** (2), 1–13.
Sachs, J.D. (2001), 'Tropical underdevelopment', NBER Working Paper 8119, National Bureau of Economic Research.
Saunderson, M.H., R.B. Haight, E.M. Peterson and R.E. Willard (1937), 'An approach to area land use planning', Land Use Planning Publication, No. 16, US Department of Agriculture, Resettlement Administration, Washington, DC.
Shleifer, A. and R. Vishny (1993), 'Corruption', *Quarterly Journal of Economics*, **108** (3), 599–617.
Tarlock, A.D. (1991), 'New water transfer restrictions: the west returns to riparianism', *Water Resources Research*, **27** (6), 987–94.
Thompson, Jr., B.H. (1993), 'Institutional perspectives on water policy and markets', *California Law Review*, **81** (3), 673–764.
Thornthwaite, C. Warren (1936), 'The Great Plains', in Carter Goodrich, W. Allin Bushrod and C. Warren Thornthwaite (eds), *Migration and Economic Opportunity: Report of the Study of Population Distribution*, Philadelphia, PA: University of Pennsylvania Press, pp. 202–50.
US Department of Agriculture (1938), *Yearbook of Agriculture, Soils and Men*, Washington, DC: Government Printing Office.
US General Land Office, *Annual Report of the Commissioner*, fiscal years 1880–1925, Washington, DC: Government Printing Office.
Wiggins, S.N. and G.D. Libecap (1985), 'Oil field unitization: contractual failure in the presence of imperfect information', *American Economic Review*, **75** (3), 376–85.
Williamson, O.E. (1979), 'Transaction-cost economics: the governance of contractual relations', *Journal of Law and Economics*, **22** (2), 233–61.
Worster, Donald (1979), *Dust Bowl: The Southern Plains in the 1930s*, New York: Oxford University Press.
Young, R.A. (1986), 'Why are there so few transactions among water users?', *American Journal of Agricultural Economics*, **68** (5), 1143–51.

PART II

PROPERTY RIGHTS AND THE LAW

6 Judicial system and property rights

Christian Barrère

Introduction

Property rights (PRs) constitute a system that defines relative rights with respect to the utilization of scarce resources, that is to say somebody's rights in relation to the rights of anybody else. As law is inefficient without law enforcement it cannot work without a judicial system. The judicial system enforces PRs by monopolizing the power of constraint that obliges everyone to accept the PR distribution and its consequences. But the judicial system plays other roles in the application of property rights. In particular, it specifies the conditions of use of property rights when there are different interpretations and when opposite claims are advanced. Hence the judicial system is a system of legitimate interpretation and distribution of the concrete effects of PRs in a social context. The judicial system has two main effects on PRs: efficiency and distribution – hence, equity. This applies to the three functions concerning the judicial system: (i) PR enforcement, (ii) PR interpretation and (iii) PR specification.

First, we shall see that the judicial enforcement of PRs is an efficient way to strengthen the incentives to cooperate and therefore to increase social welfare, especially when competition becomes tighter and when opportunism undermines the substitutes for a judiciary such as ethics or customs. Therefore judicial enforcement is a public and a club good; but as it allows the distribution of effective rights, powers and wealth to be modified, judicial enforcement does not represent a standard public good. An important consequence is that the partial and unequal character of enforcement is not only related to its judicial cost when goods are sets of characteristics. According to the distributional effects some individuals or groups may win with a strong enforcement of PRs and some others with a weak enforcement; hence, individuals, groups and organizations can develop strategies regarding each kind of PR to be enforced.

Second, PR interpretation is related to the fact that the law cannot be a perfect and complete system including a perfect property rights system, but only an imperfect property rights system. It cannot specify all the legal rights and duties in any concrete situation. The judicial interpretation is an important means for transaction cost minimization and for the formulation of rules in an open society. Moreover, we are mainly interested in an institutionalist perspective, whereby the judicial system is a producer of just decisions in front of competing interests.

Third, we study the role of the judge when specifying the concrete forms of PRs as relative rights and relative powers. Thus, the judge introduces a new process of allocation and evaluation, before and alongside the market process. The judicial decision is the last means of cutting the conflict between PR holders and those who dispute their rights. The judge organizes a system of relative concrete PRs, a configuration of relative capacities to act, that is to say a system of mutual coercion, which is similar (but preliminary) to the other system of mutual coercion, the market. In most situations, efficiency obviously plays a role, but together with equity criteria. Justice has a particular function that makes it the main connection between the various areas of society, its various dimensions, its different norms and value systems: economic, political, social, cultural, ethical and so on.

PR enforcement – an efficiency perspective

The need for an enforcement apparatus

A PR system is an efficient way to regulate the use of scarce resources because it diminishes the conflicts about uses and allows social cooperation instead of economic war. But a PR system without enforcement is no more efficient than a system without PRs. If the PR system is efficient, judicial enforcement is also efficient. The point is easy to understand in a game-theoretic presentation.

We use an evolutionary game approach because our problem is not only a problem of static equilibrium, but also of strategic choices in a context of evolving conventions. This approach makes two interesting innovations in game theory. It requires players to merely have a weak rationality. They observe the gains from the different strategies and choose the most profitable one; whereas in classical games they must conceive their strategic planning for the whole game. It introduces the possibility of random strategy moves, for instance mutations; and of strategies which become winning because of imitation. It is assumed that the players gradually prefer winning strategies. A replication mechanism associates variations in the proportion of players using one strategy and the gap between the earnings of this strategy and the average gain of the other strategies.

Let us suppose that two persons or two groups are facing PR effects (for instance, when renting a flat). The contract gives each of them some rights and some obligations. If PRs are not monitored and enforced, they tend to renege on their obligations, as is typical in free-rider problems. If we use an evolutionary game, with pure or mixed strategies (belonging to an axis from absolutely no respect (NR) to strict respect (R) of PR organization), the payoffs are ordered so that $G(nr, R) > G(r, R) > G(nr, NR) > G(r, NR)$: $\gamma > \alpha$

> δ > β; r and nr are the strategies of player (1) and R and NR those of player (2). G(nr, R) (γ) represents the temptation to cheat because the opportunistic behaviour gives the greatest payoff when the other player observes the law. And G(r, R) > G(nr, NR): α > δ because observance of law leads to cooperation, the result of which is better than a non-cooperative outcome. G(r, NR) (β) is the worst payoff, since the player is the dupe. In all cases we have the following game matrix (on the left the game and on the right the normalized game):[1]

Game [A]							
		(2)				(2)	
		R	NR			R	NR
(1)	r	α, α	β, γ	(1)	r	$\alpha - \gamma = a_1$	0
	nr	γ, β	δ, δ		nr	0	$\delta - \beta = a_2$

In the normalized game, the condition $\gamma > \alpha > \delta > \beta$ becomes $a_1 < 0$ and $a_2 > 0$. The unique evolutionary stable equilibrium is a prisoner's dilemma equilibrium, nr – NR, suboptimal. No mutant strategy can invade the game, the nr strategy being a dominant strategy (a standard replicator, with nr the proportion of players using nr, gives $dnr/dt = nr.[a_2 (1 - nr).nr - a_1(1 - nr)^2]$, always > 0).

A judicial system can be seen as a means to rule the nr strategy (no respect) out of the strategy space, so to impose a strategy of respect for the law and the associated high equilibrium r, R with the optimal outcome α, α. From an economic point of view, as Becker points out, the choice of players between r and nr strategies is rational, based on costs and benefits (including non-monetary ones, such as time or ethics). Justice cannot get rid of nr strategies (or mix strategies including nr actions), but can make them expensive. To enforce PR justice is not a purely coercive machine (all criminals are not arrested and punished, crime is not eradicated), but merely an incentive mechanism.

A partial and unequal enforcement

If players are reasonable individuals who take into account costs and benefits, enforcement of property rights needs efficient sanctions. In the general game:

Game [B]				Game [B]			
		(2)				(2)	
		R	NR			R	NR
(1)	r	α', α'	β', γ'	(1)	r	$\alpha' - \gamma' = a'_1$	0
	nr	γ', β'	δ', δ'		nr	0	$\delta' - \beta' = a'_2$

We can see that the law is only enforced in the outcome (r, R). To be the unique equilibrium of the game that implies $a'_1 > 0$, $a'_2 < 0$, so $\alpha' > \gamma'$ and $\delta' > \beta'$. Strict enforcement implies that this condition is true for all the players and under all circumstances. Observation of reality shows that this is not the case, for utility functions sometimes include a very high estimation for not respecting the law (if I am starving, stealing a piece of bread has a very high value).

The optimal sanction is also relevant (Polinsky and Shavell 2000). Infringement of PRs must be condemned not only on moral grounds, but also from an economic point of view, once the incentives created by a system of sanctions are taken into account. If infringements have unequal effects, they have to be unequally sanctioned. Thus reasonable people choose a minor infringement over a major one (when I park my car, I prefer to attack the rights of the deliveryman rather than those of handicapped people, not only for ethical reasons, but also for monetary reasons).

A third point is the cost of judicial property rights enforcement. Let us suppose that each player pays a tax T in order to finance law enforcement. We introduce this cost by substituting α'', β'', γ'', δ'' for α', β', γ', δ' (with $\alpha'' = \alpha' - T$, and so on). Then the new cooperative outcome in game [C] may be, for the individual and the collective, inferior to the old non-cooperative outcome of game [A]. It is the case if $\alpha'' < \delta$.

Game [A]				Game [C]			
		(2)				(2)	
		R	NR			R	NR
(1)	r	α, α	β, γ	(1)	r	α'', α''	β'', γ''
	nr	γ, β	δ, δ		nr	γ'', β''	δ'', δ''

Enforcement of property rights is only partial, and this is related to the judicial cost of enforcement. Enforcement is a costly operation and is all the more true as the goods are sets of characteristics. To specify and protect all PRs would require the definition and the protection of each characteristic. PR monitoring and enforcement costs are a peculiar type of transaction costs:

> Transaction costs are the costs of all resources required for transferring property rights from one economic agent to another. Transaction costs include the cost of making an exchange (i.e. discovering exchange opportunities, negotiating exchange, monitoring exchange, and enforcing it) and the cost of maintaining and protecting the institutional structure (i.e. the judiciary, police, and armed forces). (Pejovich 2001, p. xvii)

In turn, costs include three components.

The first is the choice of the social level of enforcement. If PR enforcement has a social cost, arbitration is necessary between the benefit and the cost of protection. That explains why protection is never total. Some property rights will be enforced and some will not. And the extent of enforcement changes with the cost of and the benefit from following the law.

The second relates to the configuration of the instruments. For instance the enforcement of software or music files PRs can be achieved through public procedures (courts sanction the violations of rights) or private procedures (producers introduce technical instruments as tools to protect their PRs effectively). The question is how to mix public enforcement (especially judicial enforcement) and private enforcement.

The last point is that the judicial enforcement costs are not only of a direct nature (for example, judges' salaries). To sue someone in court is costly. It implies monetary expenses and transaction costs (time, psychological costs and so on). Optimal enforcement policies have to include this problem.

Substitutes

Justice changes the earnings of r and nr strategies and, thus, increases PR observance. Some substitute means also have to be considered as producers of norms: ideology, ethics, deontology, religion, customs, culture, social conventions and so on. They can induce conventions of PR respect in place of PR violation. Similarly to when game [B] replaced game [A], they work by increasing the earnings of the r strategy in relation to the nr strategy (to exclude someone from the group for rule infraction, to reward someone for loyalty, to reprove immoral actions and so on).

These sets of norms may emerge from repeated interaction and are reinforced by it. Repetition, under different hypotheses on rational behaviours,

allows the emergence of cooperation through reciprocity or reputation building (see Hardin 1982; Fudenberg and Maskin 1990; Kreps 1990). These self-organized processes were formulated in repeated games terms (Axelrod 1984) and in evolutionary games terms (Bicchieri et al. 1997).

Following Axelrod we can imagine that cooperative strategies emerge with the repetition of the game. Nevertheless their ability to enforce property rights is doubtful. Repeated game theory has shown that strong and numerous conditions are necessary to support cooperation. Property rights will be 'spontaneously' respected in small communities (as in a family or a bureau), with identified players, frequent plays and the possibility of retaliation. It is not true when players are numerous, changing, unknown, characterized by different ideologies and ethics. Moreover, in an evolutionary perspective, the unique evolutionary stable equilibrium is the dominant non-respect strategy. If any other strategy becomes the general one (for example, the respect strategy), a player that chooses a non-respect strategy has a greater payoff and is then imitated by the others. Many examples of progressive degradation of PR structures can be recalled. Historical perspective shows that old societies based on tradition develop cooperative behaviours (for example, through gifts), depending on whether they are small, closed, stable societies, with ideological enforcement (for example, the taboo system). Whereas large, open, unstable modern societies, with many fluctuating interactions, evolving norms, refined and complex PR systems, need more explicit enforcement. In the same way the development of PR systems is accompanied by a development of judicial enforcement.

Enforcement as a public good and a club good

When enforcing PRs the judge produces a public good. Even if the sentence of a court is particular and limited, it contributes to reinforcing the weight of the PR system. It creates an incentive system that leads to cooperation and wealth. In addition to its static effect, enforcement has a dynamic effect. It increases the stability, predictability and credibility of PRs, and also provides a stable framework for expectations and investment. PR enforcement is an important means of reducing transaction costs.

The supply of public goods is characterized by free riding whenever the cost of PR enforcement is higher than individual profit and lower than collective profit. According to the dispersion of individual utilities of enforcement for some types of PRs (if I am a smoker, I am not much worried about the enforcement of anti-tobacco regulations), enforcement may vary. It becomes a club good and many clubs can coexist. History documents many types of club rules, be they local or communitarian (fishermen), or temporary (as during fairs in the Middle Ages). Today many private legal systems work and

enforce some types of PRs (for example to manage PRs on credit cards). Evolution affects the efficient configuration of these clubs, their number, types and sizes.

PRs function as a system, so that enforcement requires coherence; and the power of sanctioning deviating behaviours must be founded on legal and public legitimacy. That is why enforcement must ultimately be public.

Extension of competition

When competition intensifies, spontaneous observance of PRs becomes less likely. Let us suppose we take a symmetric game, with two equal firms; and that by following PRs a collective surplus of 10 is generated. If the two players choose the cooperative strategy (outcome r, R), each of them gains 5 (a fair share of the surplus). No respect of PRs increases the payoff (the cheater takes the whole surplus minus a cost of 2 – say, loss of reputation, cost of cheating). Thus, outcome (nr, NR) means that each player obtains a weaker result, 3 (5–2). If one is aggressive and the other is not (outcomes nr, R and r, NR), the cheater escapes his/her obligations and gains 8 (10–2), while the dupe gains none.

Game [D]			
		(2)	
		R	NR
(1)	r	5, 5	0, 8
	nr	8, 0	3, 3

The unique equilibrium, with dominant strategies, is strong but clearly suboptimal. If the firms meet regularly, the repetition of the game pushes them to cooperate, that is, to spontaneously respect PRs. So, if the repeated game can lead to the repeated Nash equilibrium (nr_T, NR_T) with (3, 3), it can also lead to the optimal repeated outcome (5, 5) with the cooperative strategies (r_T, R_T). The probability of the last one depends on the weight of the incentive to cheat: cheating immediately brings back 8 and 3 in each coming round; to go on cooperating gives 5 in each round. With ∂ as a discount parameter (the present value of a unit of utility available in the following round, which can also represent or include the probability of the end of the game in the next period), there is cheating if $3 > 2 (\partial + \partial^2 + \partial^3 + \ldots + \partial^N + \ldots)$, so for N large

enough only when $\partial < 3/5$. Rational players can therefore organize 'spontaneous' cooperation by respecting PRs, without any judicial or any other external enforcement.

In Axelrod-type situations, spontaneous respect of PRs is likely[2] because, even if their interests are not necessarily parallel, individuals are only interested in their individual payoff, and not in that of the others. They are not 'jealous'. Obviously player (1) prefers outcome (5, 3) to outcome (3, 3), but is indifferent between outcome (5, 3) and outcome (5, 5). The profit of his/her partner is strictly immaterial to him/her. This assumption is appropriate in a smooth and peaceful competition, but irrelevant in a world of tight competition in unstable markets. To gain more than my competitors can be decisive in forthcoming battles. Therefore I may prefer an outcome (2, 0) to an outcome (3, 3). However the solutions of the game are very sensitive to the levels of the payoffs.

To study this point we propose three repeated games, with the same monetary payoffs (profits) but different utility functions, corresponding to three different behavioural models, the 'egocentric alternative', the 'competitive alternative' and the 'hyper-competitive alternative'. Let us suppose that profits are monetary (3, 5, 8 billion euros) and that a level of competition expressed by the utility function $U(.)$ characterizes the models.

The first alternative, the egocentric one, is the repeated game [D] with the utility function $U_i = \pi_i$. The second one, the competitive alternative, is the repeated game [E] with the utility function $U_i = \pi_i + (\pi_i - \pi_{-i})$:

Game [E]

		(2)	
		R	NR
(1)	r	5, 5	–8, 16
	nr	16, –8	3, 3

The probability of having a cooperative outcome falls. In the game [D] it occurs with $\partial > 3/5$, now $\partial > 11/13$ is required. The hyper-competitive alternative is featured by the repeated game [F], with the utility function

$$U_i = \pi_i + (\pi_i - \pi_{-i})^2 \text{ for } \pi_i \geq \pi_{-i},\ \pi_i - (\pi_i - \pi_{-i})^2 \text{ for } \pi_i \leq \pi_{-i}.$$

Game [F]			
		(2)	
		R	NR
(1)	r	5, 5	–64, 72
	nr	72, –64	3, 3

The probability of the cooperative outcome is reduced significantly, and this equilibrium can only exist when $\partial > 67/69$. In addition, the more competitive the game, the more the discount parameter decreases, because any immediate profit is a competitive weapon that may push the competitor out of the game. Similarly, the more competitive the game, the greater the probability of reaching the end of the game, because of instability. The two elements tend to decrease the value of ∂, whereas the condition of cooperation becomes stronger.

Tight competition increases the risk incurred by cooperative players, and then increases the use of opportunistic strategies with regard to PR observance. Players fear to find themselves in the disastrous situations (r, NR) and (nr, R) and prefer the safer but suboptimal repeated outcome (nr_T, NR_T). The repetition of the game reinforces this tendency, because the 'repetition' is not in fact an identical repetition: the economic and competitive power of each player evolves, according to the results of the previous rounds. The dynamics of the game widens the differences between the profits, which are used as additional weapons in competition, and may cause the elimination of some players. The introduction of additional players would make this threat even more poignant. The use of tit-for-tat or carrot-and-stick strategies is hardly likely. In short, the extension of competition needs a more explicit and powerful PR enforcement system, which explains the growing power of judicial systems.

An evolutionary games model appears more appropriate to our purpose. Market globalization generates fewer matches between stable, well-identified pairs of players, but more between changing, unknown partners, pertaining to different cultures, characterized by unpredictable behaviours. When globalization is mature, new forms of spontaneous compliance may emerge, with a revival of cooperative behaviours. For the time being this is not the case.

This model allows us to consider a population of similar players brought to meet randomly, two by two (contrary to the previous repeated play, in which

two players and two only, played indefinitely the same game). It consolidates our former conclusions on the weak probability of spontaneous cooperative outcomes. When game [D] becomes evolutionary, the unique evolutionary stable strategy is NR. Let us suppose an invasion. A part of the group now simultaneously plays R. If they meet often enough, their surpluses in (r, R) – they obtain 5 whereas the other players obtain only 3 in (nr, NR) – can compensate their relative losses in (r, NR), where they obtain 0 against 3 with nr. Nevertheless, this never leads to an evolutionary stable equilibrium. In a configuration where every player chooses R, any deviation towards NR is winning; and the replication mechanism aligns everyone on the unique evolutionary stable strategy NR according to the dynamics:[3] $dq_{NR}/dt = 3q_{NR}(1 - q_{NR})$, where q_{NR} is the proportion of players playing the pure strategy NR. The games [E] and [F] lead to a faster alignment on the evolutionary stable strategy of PR non-observance, since the mechanism of replication becomes $dq_{NR}/dt = 11q_{NR}(1 - q_{NR})$ in the second alternative, and $dq_{NR}/dt = 67q_{NR}(1 - q_{NR})$ in the third.

PR enforcement: an equality perspective

Judicial enforcement does not represent a standard public good, since the distribution of rights affects powers and wealth. Some individuals or groups may win with a strong enforcement of PRs (the owners); and some other groups may benefit from weak enforcement (the robbers). Instead of the previous symmetrical game with parallel interests, let us now consider game [G], characterized by non-symmetrical distributions in the symmetric outcomes: the cost of the loss of reputation is always 2. In the r, R outcome, player (2) is favoured when sharing the total surplus, while in the nr, NR outcome player (1) is favoured. For instance (1) performs better in illegal conditions, and (2) in legal ones.

Game [G]

		(2)	
		R	NR
(1)	r	3, 7	0, 8
	nr	8, 0	**5, 1**

The equilibrium in dominant strategies is (nr, NR; 5, 1). Judicial enforcement reverses the result, when one player does not respect PRs, while the other respects them. That leads to game [H]:

Game [H]			
		(2)	
		R	NR
(1)	r	**3, 7**	8, 0
	nr	0, 8	5, 1

Equilibrium becomes (r, R; 3, 7). The collective surplus is increased; but while player (2) is in a better situation, player (1) is worse off. In this case, enforcement is not a common wish and needs an authority decision. Posner questions this conclusion, and argues that individuals are ready to accept a common law based on wealth maximization; in the long run they are most likely to profit from it, even if the particular application of the rule penalizes them at a given time. But of course, there can always be a permanent separation between a group with an interest in law enforcement and one against it. If so, public enforcement power becomes necessary (see the Mafia problem).

Another way to combine the distribution effects and the acceptance of regulatory control can be used within an exchange framework. In order to increase social wealth player (2) can propose player (1) to enforce PRs while modifying the shares. The non-cooperative outcome of game [G] works as a threat value (5, 1). So, every solution (r, R) with a redistribution of the type (5 + a, 1 + 2 – a) with $0 < a < 2$ is acceptable and efficient. For example, one can interpret the institutionalization of the feudal chore in these terms.

Understandably, individuals want a strong enforcement of their PRs and a weak enforcement of their obligations (opportunism). As a driver I prefer a weak enforcement of the highway code, but as a pedestrian I want strong controls. If individuals change their position within the PR system, sometimes as drivers and sometimes as pedestrians, there are minimal consequences. On the other hand, if the roles in the division of labour and the roles in the social organization are permanent and rigid, the consequences are more significant. Judicial strategies emerge and compete with one another. Being in a world of scarcity, justice can only use limited resources. The enforcement of PRs is never perfect. Justice has to decide about priorities: enforce ownership, monitor compliance with fiscal duties, enforce anti-sexism laws; weakly enforce the highway code, fiscal evasion, prohibition of prostitution, or financial delinquency.

One can also witness substitutions between modifying the law and modifying law enforcement. For instance, ‘society’ may prefer not to sanction personal

consumption of marijuana without amending the law. In some areas of newly-defined PRs (for example, tobacco regulation, sexist or racist practices, rights of handicapped people, prisoners' rights) the debate in favour of legal intervention is just beginning. The second round involves fighting to enforce these regulations. Indeed, it is not by chance that the highway code is enforced to very different degrees across countries, even in those with a similar level of development.

Judicial interpretation

PR systems not only need enforcement, but also interpretation. Justice appears as a referee among different players claiming entitlements to do something and disputing other players' rights to do the same.

The judge must connect the PR system with the facts related to a particular case. His or her first task is to draw up an assessment of the facts. The judge is a 'neutral' and 'independent' referee. He/she has to establish the 'truth', to produce the legitimate reading of the facts, whereas the parties propose different versions according to their interests.

Divergences are also related to cognitive reasons. Let us substitute the idea of limited or procedural rationality for that of substantive rationality. Individuals will then seek an average and approximative level of results, rather than the strict maximization of objective functions; they would look for 'justice' (reasonable compensation), rather than ask for maximum compensation. They will use procedures, often of an organizational nature (the individual has to decide as a member of an organization, a company, a family and so on), conventions, references to norms, thinking frameworks, 'points of view'. As a consequence, different people may build different versions of the same reality. Contemporary sociology shows that the social context influences the behaviours and leads to 'habitus', that is forms of behaviour related to the social status of the individuals. Professional litigants – say a lawyer or an insurance company – will not have the same type of behaviour as an occasional litigant. Rationalities become socio-historically situated rationalities.

At the same time the referee does not generally limit him- or herself to an 'objective' observation of facts (which corresponds more to the logic of the expert). On the contrary he/she provides a specific interpretation of them, and defines responsibilities and torts: is the damage and the PR violation intentional or not?

Hence, judges are real and active producers of interpretations. They base their findings on a specific mode of reading reality; they give more or less importance to such or such type of consideration, to social logic, psychological logic, or economic logic. Spatial comparisons show that, in the same legal area of jurisprudence, the same attacks against PRs are judged and sanctioned

in different ways. Moreover, the judge is not always a public judge. In some cases conflicting parties agree about a specific referee in order to have a specific reading of the facts. For instance firms will ask a private referee to intervene, because they expect from him/her a decision based on economic criteria, whereas they fear that a public judge would minimize them.

The law cannot be a perfect and complete system. It cannot specify all the legal rights and duties in every situation. Within the PR legal system, some PRs are very precise (shareholders' PRs), but some are vague (for example, 'human rights', privacy PRs (1985)). Williamson distinguishes three types of law, corresponding to his three general categories of economic institutions:

- the 'classical law' for the market; the law strictly specifies and records the conditions of an instantaneous market transaction. Thus, it is simple and indisputable;
- the 'neo-classical law' for the contract, a hybrid form between market and hierarchy; it organizes a longer relation than the market one does (see the industrial cooperation contract or the franchising commercial contract). It is no more than a partial framework, because nobody can anticipate all the future states of nature; and
- the 'evolutionary law' for the hierarchy; it organizes a framework to manage changes (such as labour legislation, which under some conditions allows for changes in the workers' obligations or earnings).

From an economic point of view, it appears that the legal system enforces and implements a broad spectrum of property rights. At one extreme we have precise PRs, the implementation of which has been envisaged by the parties (for instance, they signed a contract); therefore the role of the legal institution is to make them respect the contract. At the other end of the spectrum there are general rights, the concrete consequences of which are not well specified; for example civil liability specifies only general duties (to behave like a good father), without saying precisely what it implies in any circumstance. The law does not usually take into account all the specific cases but rather establishes principles, general rules. In many cases judges must interpret general principles, and then draw a solution for a particular case. The degree of interpretation is variable, larger when the application of a general PR is concerned, weaker but nevertheless not nil, when interpreting a precise PR (for example, when the contract does not envisage everything; or when it is necessary to appreciate the good will of the parties).

By interpreting the rights and the duties coming from the contracts and the PR system, justice decides on individual situations. Therefore individuals accept a self-limitation of their personal freedom in order to benefit from a public, general, and legitimate enforcement. Moreover, giving the legal appar-

atus the capacity of interpreting the law, they accept a more important abandonment of freedom. Why?

Three kinds of answer are offered by economic analysis. According to the rational choice approach, judicial interpretation derives from a rational attempt to minimize transaction costs. The second case originates from the institutional approach, where individuals are subject to a public power because they are members of a collective system – society. And society has some specific interests, which go beyond private interests. In the third case, the Austrian approach, interpretation by the judge is an organizational effect of self-evolving processes that characterize interindividual relations in an open society.

Transaction cost minimization

Transaction costs make it impossible to have a system of *complete* contracts. Therefore, the judicial system is efficient when it succeeds in reducing the social costs of a transaction through a legitimate interpretation. The specialization of the judge and the economies of scale in shaping a corps of specialized interpreters go in the same direction as the reduction in transaction costs. The extent of this reduction depends on whether they are precise or general PRs. Similarly to Williamson, Cooter and Ulen (1988) oppose the legal theory of contract, which refers to an area where transaction costs are weak, and the legal theory of liability, which covers an area where transaction costs are high. Whereas in the first case the interpretative role of justice is limited (it has to interpret the contractual wishes and agreements of the contractors), in the second case the interpretative role is much more significant. This approach also explains the absence of monopoly of the judicial system as an interpreter of the law. In the case of litigation, parties can agree on a compromise (the lawyers of the two firms reach a private agreement), or call upon a third person (a mediator). The choice of the procedures will follow the comparison of transaction costs.

Rules in an open society

In the Hayekian approach, rules are conceived as incomplete (the states of nature are not all anticipated) and imperfect (the conditions of their application can contradict one another). No rule or condition fits the infinite number of possible situations; their adaptation to a particular situation is not immediate. Friedrich von Hayek rejects any standard solution denying the specificity of the cases and the need to appreciate comparative responsibilities. So the judge is in an eminent position to interpret the PR system.

Judicial functioning results from a spontaneous self-organized process, with trials and errors, and a progressive selection. It does not consciously maximize welfare, but manages ignorance. Individuals are ignorant for three reasons:

- the future is always unknown, at least in part;
- some individuals may not know that some types of information exist; and
- some pieces of information are only partially transferable, or not transferable at all.

In this context, the judicial system allows a spontaneous disorder to become a self-organized order. Therefore, the emergence of rules, through a spontaneous self-organized process, and their judicial confirmation by the judge, allows uncertainty to be managed and reduced. Individuals can rest upon rules, which carry information and generate routines. Cooperation is enhanced by these rules, although they do not result from calculation or maximization. This guarantees their social efficiency and enables them to act as general and stable references for individual action.

The judge operates by formulating the social implicit rules on property. In other words, the main role of the judge is to make the implicit rules explicit, to clarify them, so that everyone can know them and profit from them. Consequently the judge transfers information similarly to the Walrasian 'auctioneer'.

However, the role of the judge as an interpreter is limited; being themselves under the law, judges have to comply with it. They must not set their mind on producing a new or a finalized rule, that is, an arbitrary rule. By interpreting PRs, judges reinforce their stability and make the future courts' decisions more predictable, which avoids excessive recourse to the courts.

The judicial interpretation of PR problems has a second important effect: it allows rules to evolve. The judge is confronted with problems that cannot always be solved by the existing legal systems. Pressures to modify the existing rule in the context of other laws handed down by tradition lead to changes in the rule while preserving the coherence of the legal order.

Collective organization

Three main ideas are put forward in the institutionalist perspective as regards the judicial interpretation of PRs:

- The first starts with the multiplicity of social logics in our societies. In the economic field there is a market logic and a non-market logic (hierarchical allocation of resources, cooperative systems and networks, for instance). In other fields there are other logics: communitarian or domestic (family) logics, political logics and so on. And, occasionally, these logics influence economic areas. Moreover, especially since the American and French revolutions, modern societies have been republican political societies. They define individuals as citizens and give

them political rights, independent of their market rights – for instance rights to enjoy privacy, to think, to manifest, to vote, to move (Barrère 2001).[4] The logic of these PRs is different from the logic of economic PRs. Some personal PRs are inalienable (I cannot sell my organs or my freedom or my public role). Therefore the republican logic introduces political and ethical principles, which distinguish political freedom from freedom to contract. The judge has to control the application of the principle of autonomous decision making, and prohibits 'contracts' or 'exchanges' such as 'slave–slave trader', or prostitute–upholder, minor–paedophile. In other words, the judge has to interpret the facts and the law by taking into account different principles, and has to combine individual interests and collective interest concerns.[5]

- The second point regards the abstract character of the modern legal rule. The law is based on general principles (for example, 'legal principles', 'constitutional principles') and has to be specified and interpreted in order to be applied to each particular case. The judge not only has to point out the law or the existing distribution of PRs or to apply a whole range of sanctions (for example, robbery is worth the amputation of the hand). He/she also has to produce an original solution by paying attention to individual personalities and circumstances, according to general principles.
- The third point is that the legal system, and especially the PR system, is one of *relative rights* (who has which rights, but also, whose rights prevail), of *relative capacities to act* (who can force what and up to what point), of *relative powers*, of *relative responsibilities* (what party inflicted the damage, what party should receive compensation). Hence, justice is mainly concerned with dispute resolution. It has to decide between competing interests (Mercuro and Medema 1997, p. 115). As in the Hayekian framework, the legal rules, and especially the PR system constitute together with the market the two main institutions backing relations among individuals. But the institutionalist view highlights the diversity among individuals' interests, representations and projects. If the market expresses competition, it includes a selection mechanism by the willingness to pay, although the law uses another mechanism, according to political logic and ethical criteria. Thus, a rule is not only a coordination mechanism, it is also a social compromise, an organization of relative rights and powers.

The judicial specification of PRs

It might seem strange that the judicial system is considered as a place where PRs are specified. For many economists this seems dangerous, and they refuse this kind of intervention or seek to limit it to a minimum level.

Nevertheless, judicial systems contribute to clarifying the contents and boundaries of PRs. They do this by defining individual and social norms. Their intervention is, probably, more important in European countries, where there is an old public intervention tradition, than in the United States; more important in civil law countries than in common law countries; more important in some law areas (medical liability, intellectual property, anti-competitive practices) than in others.

The definition of justice

The judge produces lawful and 'just' sentences. He/she defines not only what is legal but also what is just. The judge is not a legal scholar, but rather a creative producer; for the 'just' sentence is not necessarily already written in the law. Sometimes the judge takes a decision by choosing among many. This is obvious in criminal law. The law envisages a range of sanctions. The judge or the jury decides among these possibilities. And they adjust penance to fault according to the conditions. Sometimes, this is also true in civil law, for instance when compensation is involved. There is no standard for sanctions; no objective and indisputable amount can exactly compensate the injured parties. The judge has to make his/her own evaluation; and different judges or courts may fix different amounts. No one can use indisputable norms or conventions, since explicit markets for such damages are missing. Moreover there lacks a clear *ex ante* delimitation of who is entitled to be considered a victim. Up to what amount will the disutility of oil wastes in the ocean be compensated? And will that take into account the contaminated birds? History shows that the logic of these decisions may alter; and not only according to the changes in market prices.

Judges evaluate rights and claims; they mediate between conflicting claims and opposed PRs; they delineate the perimeter of each PR; they assess what actions are authorized by the PR and to what extent; they balance damages and compensation. Hence, their intervention in the social fabric goes deeper than just enforcing and interpreting PRs.

In other words, the judge introduces a new process of allocation and evaluation, prior to and alongside with the market process. From a PR perspective this implies two judicial functions:

- The judge specifies the precise powers, rights and obligations given to the parties by their PRs when there are opposite claims. He/she establishes who has the right to do what and under which circumstances. Has the tenant the right to require the owner to maintain the building where the flat is located? In which cases? To what extent? The judge introduces a judicial allocation of precise rights.
- To evaluate and compare rights and duties, to determine implicit prices (How much is an aggression 'worth'? How much is a damage 'worth'?

> What are their implicit 'prices'?). By doing so the judge is a substitute for the market process. He/she introduces a process of legal evaluation in a world governed by market processes of evaluation.

The judge intervenes in the interindividual relations, but does not exercise any arbitrary power. First he or she acts within the legal PR framework, under legal and constitutional principles. Second, regulatory mechanisms such as the possibility of appeal to higher courts have emerged to avoid arbitrary decisions. Yet, the use of judicial PR allocation and valuation remains to be addressed.

Miller et al. v. Schoene *and Buchanan versus Samuels*

> *Miller et al. v. Schoene* is a case which involves red cedar and apple trees and their respective owners; and cedar rust, a plant disease whose first phase is spent while the fungus resides upon its host, the chiefly ornamental red cedar tree, which is not harmed by the cedar rust. The fungus does have a severely adverse effect upon the apple tree during a second phase, attacking its leaves and fruit. The legislature of the state of Virginia in 1914 passed a statute which empowered the state entomologist to investigate and, if necessary, condemn and destroy without compensation certain red cedar trees within a two-mile radius of an apple orchard. Miller et al., plaintiffs in error in the instant case, unsuccessfully brought suit in state courts, and sued to reverse the decision of the Supreme Court of Appeals in Virginia. The arguments for the plaintiffs in error were basically simple and direct, as well as of profound heuristic value. Their main contention was that the legislature was, unconstitutionally in their view, attempting to take or destroy their property to the advantage of the apple orchard owners. (Samuels 1971, pp. 436–7; quoted by Buchanan [1972] 2001, p. 363)

Samuels's opinion According to Samuels, the court 'had to make a judgment as to which owner would be visited with injury and which protected' (Samuels 1971, pp. 438–9; quoted by Buchanan [1972] 2001, p. 363). That is, it had to choose between two conflicting claimants. For that choice the market and the exchange system must be replaced by the judicial system.

Buchanan's reply Buchanan interprets the problem in terms of externalities and proposes a solution based on the internalization of these externalities by means of exchange. If PRs are not well specified, the courts have to 'lay down the precise limits of allowable actions by the parties in question', but 'the courts are locating the limits that exist in the law; they are not, and they must not be seen to be, defining new limits or changing the pre-existing ones' (Buchanan [1972] 2001, p. 365).

Once PRs have been well specified, mutual agreement may internalize the externality as in the standard Coasian framework. It is only in the presence of 'certain narrowly-defined conditions' (for instance some free-rider obstruc-

tion) that transaction costs might give an efficiency basis for resorting to collective or state action. But, Buchanan adds, 'there is no role for judiciary' (ibid., p. 370), only for the legislative process: 'The judicial role should have been limited strictly to a determination as to the constitutionality of legislative action, and this should not have included any attempt at making a judgment as to the economic efficiency or inefficiency or to the equity or inequity of the legislative choice actually made' (ibid., p. 373).

Narrow judicial intervention as a complement of market process

As discussed earlier, judicial intervention is necessary in some situations to make rights clear. The imprecision of PRs is related to transaction costs. Specifying PRs is an essential but costly precondition for the market to operate (Papandreou 1997). In a world without uncertainty individuals, under a veil of ignorance, could *ex ante* envisage all the possible bundles of PRs, the concrete situations in which to apply them and negotiate their implementation. But with uncertainty, the cost to obtain a perfect PR system and complete contracts would be enormous. In these cases the judge can be an efficient way to specify PRs.

This reason may explain the growing role of judicial systems. For Hayek, judicial intervention is inevitable in an uncertain world (when defining land PRs in the eighteenth century nobody could imagine oil extraction), but this intervention has to be limited by its inscription in tradition.

The limits to judicial intervention

For Buchanan, the judiciary must not 'inject its own standards of value measurement in determining the constitutionality of the legislation' ([1972] 2001, p. 374). The judiciary's activism has to be rejected in favour of the respect of previously existing rights:

> There is an explicit prejudice in favour of previously existing rights, not because this structure possesses some intrinsic ethical attributes, and not because change itself is undesirable, but for the much more elementary reason that only such a prejudice offers incentives for the emergence of voluntarily negotiated settlements among the parties themselves. (ibid., p. 375)

So, he adds: 'The object of never-ending research by loosely coordinated judges acting independently is to find the law, to locate and redefine the structure of individual right, not *ab initio*, but in existing social-institutional arrangements' (Buchanan 1975, pp. 46–7, quoted by Pejovich 2001, p. xxi). Hayek, as we saw, has a similar position.

The Coasian analysis sees a conflict between claimants as a conflict about the use of a scarce resource related to an externality (often linked to technical progress, which gives more value to unexploited resources). With low trans-

action costs a precise entitlement allows an exchange of the PRs on the resource. Free exchange internalizes the externality and an efficient outcome is found. Once the PR entitlement has been well defined, the judicial system must withdraw in favour of the market. Only when transaction costs are high does the judge have to give a ruling on entitlement.

The intrusion of the judge in interindividual relations must then be limited in two ways: the market must be allowed to function (instead of setting standards of pollution, a market of pollution rights must be created); judicial evaluation must rest on the logic of free transactions, which reflect individual preferences.

Judicial intervention as an alternative to the market process

A PR specifies the limits to the use of a resource by the holder of the PR and defines all the uses prohibited to other people (I can smoke my cigarettes and my neighbour cannot steal them or cannot prevent me from smoking in my house). A perfectly specified PR would be a clear definition of all the uses that the holder is empowered to make in every state of nature. Formally, a PR system includes relations between:

- a set of title holders $\{H_i\}$, i = 1, …, n
- a set of resources $\{R_j\}$, j = 1, …, m
- a set of actions that identifies the particular uses of each resource $\{A_{jk}\}$, j = 1, …, m ; k = 1, …, z. For instance I use my car to go to work, to go on holiday, to carry luggage, to lend it to a friend, and so on.

A perfectly specified PR system implies that every action with all its consequences concerning every resource in every state of nature is unambiguously linked to the right of someone. It implies an injective application: $H_i \rightarrow A_{jk}$, $\forall_{i,j,k}$. Yet, existing PR systems are sets defined by an injective application $H_i \rightarrow R_j$, but they do not include a complete definition of the relations $R_j * A_{jk}$. No PR system can be perfect, not because of the imperfection of human nature or some kind of natural imperfection, but for economic reasons. The sets A_{jk} of all the possible uses of goods in any situation, are unknown. A perfect PR system can only exist in a world without uncertainty, in a world with complete contracts, without transaction or coordination costs. And a world without transaction costs is one with no free markets; as John Maynard Keynes said, a world without uncertainty is a cooperative economy, not a monetary and market economy. A perfect PR system also means a world where everyone has precise and undisputed powers on the resources.

In the real world, PRs are not relations between PR holders, but between holders and objects, goods, resources. In many cases the relation $R_j * A_{jk}$ is not

problematic. The application $H_i \rightarrow R_j$ is enough to define the application $H_i \rightarrow A_{jk}$ (there is no discussion whether I can lend my car to my sister instead of lending it to my brother, except if I know that she has no driving licence). This transfer from entitlement-to-goods to entitlement-to-actions is given by the law and/or conventions,[6] customs and so on. The legal principles of ownership, for instance, specify many authorized uses of an owned good, but not all of them. In some cases, there are changes in the set A_{jk} or in the effects of its elements. Then, $H_i \rightarrow R_j$ is not enough any more. The role of the law is to take into account these changes.

A judicial system has to say: 'according to your PR you are entitled to execute this action' or 'despite PRs, you are not entitled to execute it'. In many cases, A_{jk} has effects on several agents and can be related to different resources (R_j) or different holders (H_i). I am entitled to use my car in the street but if, on the same day, the New York marathon needs that route, there is a conflict. One resource, R_j (the street), is able to support opposite actions A_{jk}, marathon and car traffic. The action A_{jk} to smoke and so to pollute the air is related to different potential holders of PRs on clean air, H_i, H_j, … A PR is generally a relative power, in that it is a power for one agent and a constraint for another one. It is very difficult to organize a system of PRs, to combine everybody's rights,[7] to establish where the right of one person ends and where another's right begins. Surely, in order to respect the right to sleep one cannot make a noise at night; but what is the standard definition of 'noise'? Is it 60 decibels or 80? Is it the same in Las Vegas and in a village in Maine; in summer and in winter; on Saturdays and on other days? Attention towards children's rights is recent, but how does one combine children's and parents' rights? Combining all legal actions a priori is unimaginable, because of the transaction costs and because we are in an uncertain world.

The judge determines the boundaries for the sets of legal actions linked to PRs defined on R_j. He/she defines the relation $R_j{*}A_{jk}$. For that judgment equity criteria have to be introduced, because in most countries – if not all of them – the judge has to dispense justice, that is to apply both efficiency and equity criteria. Whenever neither the law nor the market can solve the conflict, judicial intervention is necessary. As seen in the case of *Miller et al. v. Schoene*, the judge must necessarily choose between a priori legitimate PRs: 'decide which party would have what capacity to coerce another' (Samuels 1971, p. 439). As opposed to Buchanan's argument, the prevailing PR system is not that precise. Are the landowners really entitled to plant dangerous trees, a capacity that had not been envisaged by the lawmakers?

Let us consider another interesting and more recent case. In France, in March 2001 the Court of Marseilles authorized two hundred squatters to stay on for a year. Allegedly, they had suffered from 'collective deficiencies as regards to social housing', though the occupation of a building 'without right

or title' violates the right of the ownership and constitutes 'an obviously illicit disorder' (*Le Monde*, 20 March 2001). Is that a violation of individual preferences? It does not seem so because the court pronounced a decision in a 'legal vacuum': in this particular situation neither legal text nor jurisprudence gives the answer to the question on how to combine different and opposite claims, different and opposite PRs.

From an institutionalist point of view this function of the judge has to be accepted. The judicial decision is the last means of cutting the conflict between PR holders and those who dispute their rights. The judge is the only one entitled to that function by the law. Justice is the institution that makes PRs active. Instead of leaving them in a formal state it organizes a system of relative concrete PRs, a configuration of relative capacities to act, that is to say a system of mutual coercion, which is similar (but prior) to the other system of mutual coercion, the market.

Different PRs and different logics

Judges do not apply a blueprint mechanically. They evaluate how to apply a PR system in a global legal framework and a specific context, and they have to take into account all the efficiency and equity effects. Moreover, changes in the context may modify the consequences of the existing PRs. Therefore, judges have to:

- combine the precise application of different PRs, possibly partially contradictory, and arbitrate between PRs;
- apply PRs by taking into account the specific features of the situation, of the individuals' personalities;
- apply PRs with reference to efficiency and equity.

The institutional approach claims that the efficiency criterion is not sufficient for two reasons. The first is that all PR entitlements modify relative rights, relative capacities and relative wealth. Therefore they have economic consequences: prices, production, incomes and wealth distribution are affected. The second reason is that market values, standards and norms, cannot be used as normative references. To be a 'price-taker' is to be a 'rights-taker', as if its implication as regards fairness were acceptable. Indeed, equity itself is an element in the judicial process.

We can resort to a parabola in order to clarify the relation between entitlements and solutions with distribution effects. I called this the 'Titanic model'. A shipwreck may create a scarcity of lifeboats, if there are not enough to accommodate all passengers and crew. We are here in a typical situation with ill-defined property rights on a resource that is becoming tragically scarce. We have to specify entitlements. The efficiency criterion is inoperative: what-

ever the PR allocation, no more lifeboats are to be constructed. A wild conception of efficiency – give the seats to the rich – is questionable from an ethical point of view. A naive conception of efficiency – give the seats to those with the highest preferences – makes little sense. Surely, entitlements are a problem of equity. Many solutions can be envisaged (and different conventions can exist). Priority can be given to women and children, according to a lottery, age, gender, hair colour, political preferences, nationality, physical strength and so on. This shows the 'arbitrariness of the rights'. As in the critical philosophy and sociology of Pierre Bourdieu and Michel Foucault, every right is related to primary entitlements through free exchanges or power relations. These entitlements are not the legal translation of some pre-existing fairness, efficiency or equity, but a part of a specific system of PRs among many potential systems. In most situations efficiency has obviously to play a role but together with equity criteria. In the Coasian framework, when there are no (or few) transaction costs, entitlements are neutral for efficiency, so that the judge may favour equity criterion (in the standard example of air pollution, he/she assigns a right to clean air, which means that he/she decides who pays for the anti-pollution filter). When there are transaction costs and the law does not establish priorities, the question is how to combine different types of criteria that lead to different outcomes. In these cases, the judge has to evaluate case by case.

Judicial intervention is all the more necessary when the market is unlikely to produce equitable results.[8] Then, if the judge only defines PR assignments with reference to market prices and rights, it turns out that market data themselves derive from the system of pre-established rights,[9] which are disputable or even disputed. The market cannot constitute the fixed point since it requires another point of reference, the legal one. This is why some authors insist on the need for judicial transparency when fairness is at stake.

The growing role of the judiciary

The growing role of the judicial system is correlated to more complex PRs and more complex uses of economic resources, with fewer free goods or resources. It is related to increasingly sophisticated economic regulation. A second reason is the rise in the number of opportunistic strategies as a consequence of competition.[10] A third reason is that judicial decisions must follow different logical patterns. Justice has a particular function that makes it the main point of connection between the various areas of society, its various dimensions, its different norms and value systems: economic, political, social, cultural, ethical and so on. At the same time it is related to the refusal of absolute constructivism, which would lead to a perfect property rights system, with no inconsistencies, no vacuum and no contradictions. Rights are, in some proportion, necessarily incoherent. They often have dif-

ferent sources, origins; they are linked to different legal principles. Surely, they refer to different value systems, not necessarily of an economic nature, or exclusively generated by the market process.

Notes

1. The symmetric structure of the game is not a restrictive hypothesis. The payoffs are utility indices and may be chosen arbitrarily as being equal in the same outcomes for each player. The important point is that, for everyone, $\gamma > \alpha > \delta > \beta$.
2. Optimal cooperative equilibrium is possible, but the possible outcomes are infinite. Any system of identical strategies (to cooperate *n* times, not to cooperate *m* times, to cooperate *n* times …) constitutes a system of best responses and is a solution, according to the folk theorem, with the two extreme systems $n = 0$ (no-cooperation indefinitely repeated with 3, 3) and $m = 0$ (cooperation indefinitely repeated with 5, 5).
3. $dq_{NR} / dt = q_{NR} . [G(nr, S) – G(S, S)] = 3\, q_{NR} (1 – q_{NR})$, with q_{NR} the proportion of players using the pure strategy NR, S being every strategy in kind $[pS1, (1 – p)S2]$, $0 \leq p \leq 1$.
4. The idea of the autonomy of the political logic of law is sustained in the United States by the 'modern civic republican tradition school'; see Mercuro and Medema (1997, p. 97).
5. Posner himself notes 'the need for specifying an initial set of individual entitlements or rights as a necessary prerequisite for operationalizing wealth maximization' (Posner and Parisi 1997, p. xi), and writes that it is difficult to define the absolute and relative importance (according to other components of justice) of efficiency research in justice research. He therefore accepts the analysis by Calabresi, according to whom justice is not an objective or a criterion like efficiency or distribution, and cannot then be the object of a trade-off with them. Efficiency and modes of distribution are only elements, 'ingredients' of justice, which remains a goal of a different nature. Posner adds that efficiency is congruent with our moral intuitions because it is founded on assent. Efficiency based on exchange respects personal freedom; corresponds to our search for improvement (Posner 1983, p. 89, quoted by Mercuro and Medema 1997, p. 60).
6. Am I entitled to smoke in my neighbour's house if he hates smoke? There is a potential (even unimportant) conflict between my individual freedom to smoke and his individual freedom to breathe fresh air. The solution is conventional (to be polite …).
7. To enforce ownership is the same as prohibiting a hypothetical right to steal. Similarly, monopolistic rights imply exclusion (except for public goods). Prohibition is the other face of freedom.
8. Moreover, unequal positions in the market may have a cumulative effect. See the debates on free trade and protectionism in developing countries.
9. 'The economy is a system of power, of mutual coercion, of reciprocal capacity to receive income and/or to shift injury – whose pattern or structure and consequences are at least partially a function of law' (Samuels 1971, p. 440).
10. See the *Prestige* tanker shipwreck on November 2002 off the Spanish coast of Cape Finisterre, and its incredible maze of overlapping PRs, for an example of the opportunistic strategies and PR complexity.

References

Axelrod, Robert (1984), *The Evolution of Cooperation*, New York: Basic Books.

Barrère, Ch. (2001), 'Marchandisation et judiciarisation' [Development of Market Organization and Development of Judicial Intervention], *Economie Appliquée*, **54** (3), September, 9–37.

Bicchieri, Cristina, Richard Jeffrey and Brian Skyrms (1997), *The Dynamics of Norms*, New York: Cambridge University Press.

Buchanan, J.M. (1972), 'Politics, property, and the law: an alternative interpretation of *Miller et al. v. Schoene*', *Journal of Law and Economics*, **15** (2), October, 439–52; reprinted in Svetozar Pejovich (ed.) (2001), *The Economics of Property Rights*, vol. I: *Cultural, Legal and Philosophical Issues*, Cheltenham, UK and Brookfield, US: Edward Elgar, pp. 362–75.

Buchanan, J.M. (1975), *Freedom in Constitutional Contract*, College Station, TX: Texas A&M University Press.
Cooter, Robert and Thomas Ulen (1988), *Law and Economics*, New York: Harper Collins.
Fudenberg, D. and E. Maskin (1990), 'Evolution and cooperation in noisy repeated games, *American Economic Review*, **80**, 274–9.
Hardin, Russel (1982), *Collective Action*, Baltimore, MD: Johns Hopkins University Press.
Kreps, David M. (1990), *A Course in Microeconomic Theory*, New York: Harvester Wheatsheaf.
Mercuro, Nicholas and Steven G. Medema (1997), *Economics and the Law, from Posner to Post-Modernism*, Princeton, NJ: Princeton University Press.
Papandreou, Andreas A. (1997), *Externality and Institutions*, Oxford: Oxford University Press.
Pejovich, Svetozar (ed.) (2001), *The Economics of Property Rights*, vol. I: *Cultural, Legal and Philosophical Issues*, vol. II: *Property Rights and Economic Performance*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Polinsky, A.M. and S. Shavell (2000), 'The economic theory of public enforcement of law', *Journal of Economic Literature*, **38**, March, 45–76.
Posner, Richard A. (1983), *The Economics of Justice*, Cambridge, MA and London: Harvard University Press,.
Posner, Richard A. and Francesco Parisi (1997), *Law and Economics*, vol. I: *Theoretical and Methodological Foundations*, Cheltenham, UK and Lyme, USA: Edward Elgar.
Samuels, W.J. (1971), 'Interrelations between legal and economic processes', *Journal of Law and Economics*, **14** (2), October, 435–50.
Williamson, Oliver E. (1985), *The Economic Institutions of Capitalism*, New York: The Free Press, Macmillan.

7 On the coexistence of different property rights systems – and its consequences for economic growth and development

*Stefan Voigt**

Introduction

The Hindu tradition of burning a widow on the funeral pyre of her dead husband was officially abolished in British India in 1829. Yet, instances of 'suttee' – as this tradition is also called – still occur today. Also in India, although a dowry is officially prohibited, advertisements in the classified sections of many newspapers make little effort to disguise the fact that a substantial dowry is expected.

Many states the world over have tried to ban the consumption of alcohol by outlawing its manufacture, transportation or sale. The United States even changed its constitution, the eighteenth amendment establishing prohibition (and the twenty-first repealing it). Enormous fortunes were made as a result of prohibition, Al Capone probably being the most famous figure of the time. The prohibition of trading some goods and services has often had very similar effects; just think of prostitution or the drugs trade.

In many 'multicultural' societies, property rights systems partially overlap: is a Muslim teacher allowed to wear her head-shawl while teaching in a German school? Do Sikhs have to wear a helmet while riding a motorbike? Are Muslims – and Jews – allowed to kill animals according to their tradition although this way of killing animals is generally prohibited? How are German courts to decide whether fathers of Turkish or Kurdish girls can have their daughters killed because their behaviour has brought disgrace on the entire family? Should they apply strict German standards or should they take the cultural context into account?

These examples seem very heterogeneous. We claim that in all of them, the coexistence of different property rights systems plays a crucial role. Here, we are mainly interested in the economic consequences of such property rights systems. It is the main hypothesis of this contribution that the coexistence of different systems is not automatically a hindrance to economic growth and development – as is often argued. But if the coexistence of different systems can be both beneficial as well as a hindrance to economic development, then one needs to spell out the conditions under which one or the other will be the case. Before we can turn to the economic consequences, a number of ques-

tions need to be dealt with, namely: (i) How do we define property rights systems? (ii) What does coexistence of different property rights systems mean? (iii) How does coexistence emerge and how relevant is it empirically? We then turn to the economic consequences of the coexistence of conflicting property rights systems, and examine the relevance of values and norms for property rights systems to be enforceable. In the penultimate section, a somewhat competing hypothesis is examined, namely that it is the legal history of a country that determines its future prospects for economic growth and development. This hypothesis has received some clout in the past couple of years. The last section concludes.

How do we define property rights systems?

One of the founding fathers of the property rights approach, Svetozar Pejovich (2001, p. xiii) has defined property rights as 'relations among men that arise from the existence of scarce goods and pertain to their use'. A property rights *system* would then be a coherently ordered number of property rights that cover a potentially very high number of relations among people and regulate the use of scarce goods.

This definition implies that property rights are not confined to relationships established by means of private law (a contract concerning the exchange of a good being the paradigmatic example) but also include relationships regulated via public law: the right to vote is explicitly mentioned by Pejovich (ibid.). Property rights systems thus pertain to both private and public law. Many legal scholars would argue that in order to classify as property rights, individuals should have the title to make the state enforce their rights. Such a position is not equivalent to the position of legal positivism as conventionally defined. A legal positivist would argue that all rules generated by a certain procedure are legitimate law. Legal positivism is thus concerned with the generation of rules. John Austin ([1832] 1977), for example, defined positive law as the commands of a sovereign addressed to political inferiors. The legal scholar who demands of a property right that it should be enforceable by representatives of the state is not focusing on the process of producing law but on the aspect of its enforcement. Most of the time, both aspects are stressed by legal positivists but logically, this does not need to be the case. One could call this point of view 'state positivism' because it presupposes the existence of the state and some state enforcement agencies in order to talk of the existence of property rights.

It is no doubt true that the state has become extremely important in enforcing property rights. Yet, the probability that different property rights systems coexist that are all created and enforced by the state seems to be rather low. And, as the examples presented in the introduction demonstrate, the behaviour of many individuals is not only structured by property rights that are

enforced by the state but also by rules that are enforced by other individuals or organizations. We thus propose a broad delineation of property rights that is independent from the state as its role of legislator as well as that of adjudicator. There might be property rights that are neither created nor enforced by the state. Economics as a social science is interested in explaining the behaviour of individuals based on their preferences and restrictions. As long as the possibility exists that behaviour is channelled by rules that are neither created nor enforced by representatives of the state, this can have far-reaching effects on economic growth and development.

Rules that are supported by an enforcement mechanism we shall call 'institutions'. Further, we propose classifying institutions with regard to the kind of enforcement mechanism used. A general dichotomy could differentiate between external institutions which are backed by the coercive monopoly of the state and internal institutions which rely on private enforcement or enforcement internal to society, hence their name. Among five types of institution, four types of internal institution can be distinguished according to the enforcement mechanism used (see Table 7.1).

Table 7.1 Types of institution

Kind of rule	Kind of enforcement	Type of institution
1. Convention	Self-enforcing	Type 1 internal
2. Ethical rule	Self-commitment of the actor	Type 2 internal
3. Customs	Via informal societal control	Type 3 internal
4. Private rule	Organized private enforcement	Type 4 internal
5. State law	Organized state enforcement	External

Conventions are solutions to pure coordination games, in which all participants are better off if they coordinate their behaviour. There is no conflictual element, so no participant has a preference for any particular Nash equilibrium in the event that there is more than one. Once a particular equilibrium, a convention, has emerged nobody is able to make him-/herself better off by deviating unilaterally from it. It is thus self-enforcing. Conventions are 'type 1 institutions'. The rule that everybody should be driving on the right-hand side (left-hand side) of the road is probably the most famous example for this type of institution. Since no individual can make him-/herself better off by unilaterally deviating from the convention, state enforcement is, at least in principle, not needed. Conventions fit into the definition of property rights cited above as they define the – conventionally accepted – use of the road which functions as the scarce good.

Moral norms are defined as prescriptions on how to behave (or how not to behave) in specific interaction situations irrespective of the consequences on the utility of the involved individuals. Their observance can be secured either by internalization or by private and spontaneous third-party enforcement. Moral norms, whose enforcement is secured by the actor him-/himself, that is by internalization, are 'type 2 institutions', whereas those enforced by others are 'type 3 institutions'. The latter implies an unspecified variety of individuals monitoring the compliance by way of spontaneous control, one possible example being to sanction non-compliance by informing others about this behaviour in order to diminish the reputation of the person who did not comply. If the threat of losing one's reputation is powerful enough to channel the behaviour of individuals, rules that are enforced like that can have wide-ranging consequences. Just think of the father of the Kurdish girl who prefers to kill his own daughter rather than lose his reputation as the father of a respected family.

The fourth type of internal institution involves rules whose enforcement is secured by some kind of private organization. Enforcement may rely on private courts of arbitration that monitor compliance. The enforcement of rules by private organizations is a 'type 4 institution'. Examples for private organizations that enforce property rights are (international) chambers of commerce but also religious associations.

Rules whose non-compliance is sanctioned by the state, that is, a very specific organization, are 'external institutions' because the sanctioning is external to society. Laws and decrees are examples for external institutions.

If the existence of property rights systems is not dependent on their being enforced by the state, the possibility arises for a large variety of different property rights systems to coexist.

What does coexistence of different property rights systems mean?

In the last section, we argued that there are a variety of ways in which rules can be enforced. If there are various enforcement mechanisms, they could be used to enforce different rules. As soon as that possibility is acknowledged, possible relationships between the various kinds of institution just introduced move to centre-stage.[1] Four kinds of relationships can be distinguished: (i) they can be *neutral*, that is, the institutions regulate different areas of human interaction; (ii) they can be *complementary*, that is, the institutions constrain human behaviour in an identical or similar fashion and rule-breaking behaviour is sanctioned by private individuals as well as by representatives of the state; (iii) they can be *substitutive*, that is, the institutions constrain human behaviour in a similar fashion, but rule-breaking behaviour is sanctioned *either* by private individuals *or* by representatives of the state; and (iv) they can be *conflicting*, that is, the institutions constrain human behaviour in

different ways. Abiding by an internal institution would then be equivalent to breaking an external one and vice versa. If the relationship between internal and external institutions is conflicting, this will increase coordination costs. If it is complementary, it will decrease them, because the state has to provide fewer resources for the enforcement of its institutions.

We shall deal exclusively with conflicting relationships here as – at least prima facie – the economic consequences would appear to be the most problematic: if, say, a traditional property rights system administered by the local mullahs is respected, this is equivalent to violating the system that representatives of the state are trying to enforce on a given population. Generally, one could talk of the coexistence of different property rights systems if more than one such system is in place at one location at a single point of time. Here, this is further reduced by talking only of the coexistence of different property rights systems if the respect of one such system is equivalent to violating another one that exists at the same location at the same point of time.

In principle, all four kinds of internal institution can be in conflict with state-enforced external institutions. Even though a convention is a solution to a game with at least two Nash equilibria, the government's attempt to switch from one to another might arouse opposition. People could not only have become accustomed to a particular convention, but also have invested sunk costs, for example, installing a right-hand drive on a car. Switching to a new convention would thus be costly. Ethical rules are often based on religious concepts. If new rulers follow a different faith, actors will often remain committed to their traditional beliefs – and act accordingly. Customs can conflict with newly introduced property rights. The same holds true for the rules administered by organizations such as trade guilds.

How does coexistence emerge and how relevant is it empirically?

The answer to these questions could be another one namely, why should there be only one property rights system? It is often forgotten that the notion of the (nation-)state whose representatives demand complete control over the rules as well as their enforcement is a rather novel concept and that over long periods of time in history, the coexistence of conflicting property rights systems seems to have been the norm. But given that we accept the notion of state whose representatives claim to have the monopoly of legitimately threatening coercion, what are the reasons for the coexistence of conflicting property rights systems?

Property rights define the wealth of a person as well as possible income streams. It is thus very likely that their definition is subject to quarrels induced by conflicting interests concerning their exact delineation. If one individual, or, rather more likely, a group of individuals, is very powerful

compared to the rest of the respective society, this group has the chance of defining property rights in such a way as to maximize the expected utility of its members.[2] Societies dominated by very powerful groups are thus not expected to experience the coexistence of conflicting property rights systems. If, however, none of the groups quarrelling on the precise definition of property rights systems clearly dominates all the other groups, the emergence of conflicting property rights systems is a possibility.

The absence of a dominant group is thus a necessary condition for the emergence of conflicting property rights systems. It is, however, not a sufficient condition. A variety of systems will only emerge if those members of society who are not a member of the most powerful group are able to coordinate their interactions according to some alternative property rights scheme. Traditions based on internalized values and norms are an obvious candidate here as they will often have the quality of being focal points (Schelling 1960). Religious leaders might at times be well suited to offering alternatives to the property rights system supplied by a powerful group. If an individual has a genuine choice of which property rights system to follow, the coexistence of systems promises a higher degree of individual liberty and it might even have positive effects. This topic will be taken up in the next section.

The discussion on the necessary and sufficient conditions for the emergence of conflicting property rights systems has been somewhat abstract, so we shall add a number of examples:

- First, suppose a state has just lost a war and the foreign winner has an interest in modifying the property rights systems currently in place. Although it has just won a war, it might not be powerful enough to implement its property systems completely and many individuals in the state that lost the war might continue to structure their interactions using the property rights system that they are used to.
- Second, a very similar situation can arise between colonizer and colony: the colonizer might be powerful enough to form a new colony but not dominant enough to implement its own property rights system.
- Third, suppose there has been a radical change of government domestically, for example a revolution, a coup or an attempt to install a democratic government. Revolutionaries might try to use the property rights system as an instrument for social change. However, if they are not powerful enough and substantial parts of the population prefer a different property rights system, this might give rise to the coexistence of conflicting systems. This difficulty can also arise when a new government tries to implement a property rights system adequate for the establishment of a market economy. If there are important groups who expect to lose due to its establishment, they might prefer to coordinate

their interactions using different rules. Some of the transition processes of the last decade provide ample evidence for this phenomenon.[3]

- Fourth, societies are multicultural. If the various groups hold different values and norms, their traditional ways of structuring interactions might appear more attractive to many group members than the property rights system offered by the state.
- Finally, attempts to harmonize property rights systems on an international level might alienate individuals from the newly created international systems and make them stick to their national ones. This factor may become more important in the future.

We have given reasons for the possible emergence of conflicting property rights systems and provided some examples; now we shall turn to the question of how such systems can be ascertained empirically. This is, quite obviously, no mean feat. Relevant aspects include (i) the identification of the different systems, (ii) their difference concerning substance as well as procedure and (iii) their relative relevance. Note that this is not an evaluation of the economic consequences, but these steps must precede such an evaluation. Ostrom (1996, p. 208) writes: 'These rules may be almost invisible to outsiders, especially when they are well accepted by participants who do not even see them as noteworthy'. The 'difference' between the various systems can only be measured if one has a generally agreed-upon metric at one's disposal, which is not the case.

Ascertaining their relative relevance could be done by counting how many transactions are conducted according to which property rights system. Moreover, one could also count the number of times that various conflict-settling mechanisms are used. No empirical measure to ascertain conflicting property rights systems is readily available. However, we shall now discuss two possible proxies.

First, we could try to estimate the importance of the shadow economy or the informal sector in a given country. Almost all transactions need some sort of structuring mechanism. Transactions executed without resort to either state-created law or state-run enforcement agencies can be thought of as being structured, drawing on some alternative property rights system. The size of the informal sector can be hypothesized to be a good indicator of the relative quality of the property rights systems provided by the state: the costlier it is to structure interactions using that system, the larger the proportion of interactions that will be secured by drawing on alternative systems. It is important to bear in mind that the size of the informal sector is a measure of *relative* quality: it says nothing about the quality of state-offered property rights (or the alternative property rights system) *per se*, but only about their relationship. Furthermore, it is important not to assume implicitly a fixed

number of interactions: if neither the state-run property rights system nor any of the alternative systems are capable of reducing uncertainty considerably, the total number of transactions taking place will be lower than if either of the systems performed satisfactorily.

Second, corruption levels could also be used as a proxy for the relevance of conflicting property rights systems. Corruption has been defined as 'the misuse of entrusted power for private benefit' (Transparency International 2000, p. 1). For our purposes here, we shall interpret bribes as the price for switching from a state-run property rights system to an alternative one. The amount of bribes paid could then serve as an indicator for the relevance of conflicting systems.

The purpose of paying bribes is to obtain results that are different from those that would have been obtained had one not paid the bribe. Bribe paying and corruption are themselves based on informal rule systems.[4] The problem with this proxy is that it does not contain any information on the alternative property rights system(s) itself. Additional problems are (i) no objective data are available[5]and (ii) the readiness to pay a price for not using the state-run system implies that this system does have some relevance. Assuming that the various state-run systems had the same relevance would be a heroic assumption.[6] In general, one could still expect corruption levels to be determined by the adequacy of property rights systems as factually implemented.

Even though we currently do not have an indicator at our disposal that would enable us to measure and compare the degree to which various societies experience conflicting property rights systems, there are a number of case studies that would seem to suggest that conflicting systems have always been a problem: they were a problem before the arrival of the nation-state and they have remained so even after the nation-state had been firmly rooted in many parts of the world. The examples given in the introduction, many of which are current, can all be interpreted as cases of conflicting property rights.[7]

What are the economic consequences of conflicting property rights systems?

Relying on the plausible assumption that conflicting property rights are an empirically relevant phenomenon, we now want to ask whether they also are an empirically relevant problem in the sense that they inhibit growth and development. As the empirical assessment of conflicting systems is difficult as such, estimates concerning their effects on economic growth will be equally difficult. In this section, some possible benefit and some cost components will be discussed. As comparable data are lacking, some case studies will be examined.

A first possible cost component of conflicting property rights systems is the cost of familiarization with more than one system, thus increasing transaction costs.

The coexistence of conflicting property rights systems means that if an individual behaves in accordance with one system, he/she reneges upon another, coexisting system. In order to escape sanctions, it will often be necessary to execute transactions in a clandestine way which will lead to suboptimal firm size, the omission of advertisement and marketing campaigns, the impossibility of financing certain projects externally and so on. This is thus a second kind of transaction cost. In general, high transaction costs will lead to fewer transactions taking place, the division of labour will be less, and economic growth and development are lower than they would be with an efficient property rights system.

Suppose that the representatives of the state are keen on having a monopoly in the supply of a property rights system. If there is some alternative, representatives of the state will have to spend more resources on implementing their system. In part, these will be made up of additional monitoring costs that are incurred to ensure that alternative systems are not used. If there were no conflicting systems, resources could be put to a more productive use.

Property rights systems can be called efficient if they enable the actors to execute transactions at low cost and if non-compensated externalities between actors are absent. Suppose such efficiency were possible. Under that assumption, the existence of more than one system must by necessity imply that $(n - 1)$ other systems are inefficient. The coexistence of conflictual systems would thus be equivalent to welfare losses.

The question is whether such a delineation of efficiency is empirically meaningful. The underlying notion of a social welfare function has often been criticized; the critique does not need to be repeated here, but we should point out some problems with regard to the issue at stake. The definition of 'externalities' depends on a person's values and norms. If that is true, then different values and norms could lead to the necessity to internalize the effects of quite different behaviour. The attempt to delineate one universally efficient property rights system would thus be in vain.

Some case studies

Let us now turn to some case studies. The most important empirical study in this line of thinking is de Soto's (1990) study of three informal sectors within the Peruvian economy: informal housing, informal trade and informal transportation. De Soto's central conjecture is that the size of the informal sector is a function of the compatibility of state-run property rights and the values and norms of the concerned parts of the population, or, in his own words (1990, p. 12): 'We can say that informal activities burgeon when the legal system imposes rules which exceed the socially accepted legal framework – does not honor the expectations, choices and preferences of those whom it does not admit within its framework – and when the state does not have

sufficient coercive authority'. Although the informal sector is far from anarchic, de Soto is very careful not to glorify its performance. With regard to informal business, he notes that most businesses will have to forgo scale effects because beyond a certain size, it will be impossible to remain informal, that they will often remain undercapitalized because they cannot provide the banks with the necessary securities, that they will be excluded from using certain markets such as stock markets and trade fairs, and that transactions will be accompanied by substantial information costs. Furthermore, long-term investment might well be impossible, which means that the investment rate in the informal sector will *ceteris paribus* be lower than that of the formal sector.

Ellickson (1986, 1991) wondered whether Coase's (1960) famous example of the farmer and the rancher had any empirical content. Coase maintained that conflicts between farmers, on the one hand, and ranchers whose cattle trespass on the farmers' land, on the other, are decided by the legal system. Ellickson (1994, p. 97) maintains that this view 'is almost certainly incorrect in any rural area in which neighbours repeatedly interact'. Ellickson studied how disputes between ranchers and farmers were settled in Shasta County, California. This county was chosen because in some parts of it ranchers are strictly liable for cattle trespassing, while in others they are not. He was able to show that, no matter what legal rules prevailed, the way neighbours resolved their disputes remained unchanged, that is, the property rights system run by the state did not have any effect on the kind of conflict resolution chosen. This study tells us that, under certain circumstances (repeat interactions), privately administered property rights systems still trump state-run ones, even in such highly developed societies as those in California.

We have already noted a number of cost components applicable in cases of the coexistence of conflicting property rights systems that can be detrimental to economic growth and development. But would farmers in Shasta County really be better off if they had only one (supposedly state) law to turn to? Phrased in a more general way: under what conditions is the coexistence of property rights systems conducive to additional wealth – and under what conditions does it prevent economic growth? To rephrase the question: if there is competition between the coexisting systems, under what conditions will it lead to higher growth and under what conditions will it prevent growth? We now turn to possible benefits of the coexistence of conflicting property rights.

The competition of systems

The competition of (property rights) systems and its welfare effects have been debated ever since Tiebout's (1956) seminal paper which concerned actors moving to the municipalities that supplied public goods bundles which

best suited their preferences. The paper proved to be the starting point for discussions concerning the effects of institutional competition between member states of federations as well as between nation-states. Critics have claimed that such competition could lead to a 'race to the bottom' (for example, Sinn 1997); adherents believe it to be an ingenious mechanism to get rid of inadequate property rights systems. This lengthy debate cannot be surveyed properly here (however, see Kiwit and Voigt 1998; Voigt 1999a, ch. 8). It is closely related to the issue discussed here, yet not quite identical: the debate triggered off by Tiebout usually centres around the competition of property rights systems provided by nation-states; here, however, we are interested in the competition of such systems supplied by the state with those not supplied by the state.

Competition in general can be divided into (i) an exchange process taking place between the supply and the demand side and (ii) a parallel process in which the actors on the same side of the market observe one another's behaviour. Competitors can imitate successful innovations, or they can try to top successful innovations by innovations of their own. As long as they perceive themselves as acting in the same market, their own success also depends on the behaviour of their competitors. A first condition that needs to be fulfilled if the competition of property rights systems is to have beneficial effects therefore is that suppliers perceive themselves as competing with other suppliers. If they do not care what system individuals turn to, it is unlikely that a particular system will be improved because a lower number of transactions are structured according to its rules. The suppliers must enjoy some utility from individuals using 'their' system.[8]

It was pointed out above that a variety of property rights systems means higher transaction costs. Competition between systems will only be beneficial if it leads to an improvement in rules that overcompensate for the additional transaction costs that have to be incurred in order to become familiar with the competing systems. But that is close to a tautological formulation.

Viktor Vanberg (1992, p. 111) has described the preconditions that are required for institutional competition to function effectively, namely '(a) whether potentially wealth-enhancing innovations can be and are likely to be tried out; and (b) whether the mechanism of selective retention reliably operates as an error-eliminating mechanism, that is, that it reliably selects against less efficient practices (tools, routines) and for wealth-increasing practices'. These preconditions mean that wealth-enhancing institutional innovations can be identified as such. One might therefore be dealing with a problem of identification and/or evaluation. For anybody sceptical of interpersonal utility comparisons, this can constitute a serious problem.

If the notion of institutional competition implies the hypothesis that the 'better' institutions will prevail, one needs a criterion to make institutions

comparable. In economics, it seems justified to take per capita income as the most general criterion. 'Better' institutions would thus translate into higher per capita income. But a high per capita income would be the consequence of an entire set of institutions and not any single one. Supposedly, there is no unequivocal mapping between institutions and per capita income.[9]

These preconditions also mean that there must be some mechanism which ensures that those institutions that are (comparatively) less wealth enhancing are eliminated. To check whether this is generally the case with regard to the competition of property rights systems, we should be able to specify a mechanism of systematic elimination. Often, this will not exist as different groups of actors have different concepts of 'efficiency'. But let us look at a process that is often described as a success story.

The legal development of Europe is often said to owe its success precisely to the coexistence of conflicting property rights systems (see, for example, Berman [1983] 1991; Jones 1987, p. 112). Berman identifies ten characteristics of what he calls the Western law tradition. The ninth of these characteristics is the coexistence and competition of various jurisdictions and legal systems within a single society, which, according to Berman, might be the single most important characteristic of the Western law tradition. This pluralism was, or used to be conducive to economic growth. Similarly, it was, or used to be, a source of freedom. Berman prefers the past tense here because he believes that attempts were made in the twentieth century to rationalize the various jurisdictions and legal systems into one centralized legislation and administration.

What were the coexisting property rights systems and how far did their jurisdiction extend? The most basic distinction is between canonical law on the one hand and non-canonical law on the other. Non-canonical law was not, however, a unified body of law – or a homogeneous property rights system – but consisted of a variety of systems that were conflicting at least in part. Berman explicitly deals with feudal law, manorial law, the law merchant, city law and king's law. He further claims that everybody in the Western Christianity ambit lived under at least two systems of law. Berman calls the entirety of legal systems 'legal order'.

In general, he distinguishes between two kinds of jurisdiction – that over certain groups of people and that over certain behaviour. The Church claimed to have jurisdiction over: ecclesiastics and the members of their households; students; crusaders; the poor, widows and orphans; Jews in cases in which Christians were also involved; and travellers, including merchants and sailors. The Church claimed jurisdiction over behaviour involving: the sacraments; testaments; prebends; oaths; and sins that were supposed to be sanctioned by the Church. In the twelfth century, the emerging legal science developed different areas of material law out of these groups, namely (i) family law, (ii)

law of succession, (iii) property law, (iv) contract law, and (v) criminal and tort law.

Berman believes that the plurality of legal systems as well as the plurality of courts increased freedom; in fact he evaluates the competing legal systems within one geographical area as an early form of the separation of powers ([1983] 1991, p. 471). Translated into our terminology, Berman describes the competition between the suppliers of *internal* institutions, namely those institutions that are administered by an organization without being backed by the state in Max Weber's ([1920] 1988) sense. It is exactly this sort of competition that has been identified by various historians as the explanation for the high rates of economic growth that Europe experienced in comparison to other regions of the world (Jones 1987; Rosenberg and Birdzell 1986; see also Kennedy 1987). There are economists who claim the competition between *external* institutions to be advantageous. They are often quick to quote these historians as providing evidence in favour of that claim without, however, paying due attention to the difference between external and internal institutions. This is problematic because nation-states systematically suppress competition of the kind described by Berman. He has thus chosen his tense with care. The emergence of the nation-state soon led to the suppression of this kind of institutional competition. The number of potential solutions to coordination problems that could be tested simultaneously thus decreased dramatically. One should therefore also expect growth rates to have decreased. This was, however, not the case.[10]

Non-state-run property rights systems are often based on internal institutions such as ethical rules or customs, which are not prone to deliberate modification. If the state-run system and an alternative system based on deeply ingrained values and norms are not in accordance with each other, the external property rights will not bring about the desired effects. Even worse, the system could soon become a dead letter. If a close correlation between *de iure* and *de facto* property rights systems is normatively striven for, a policy implication can be derived: if values and norms are deeply ingrained in society and are largely exempt from deliberate modification and the coexistence of property rights systems is a barrier to growth and development, then state-administered systems should not be fundamentally at odds with the valid internal institutions of a society.

Now, it could be countered that many societies might have grossly inefficient systems at their disposal and that this was a major reason for their relative deprivation (North 1990). Reaching a correct diagnosis is, however, not the same as having a ready-made therapy at hand. Societies might indeed be constrained by their values and norms, and improvement might only be achievable in very small steps or – in other words – path dependence might be a hard constraint. In the (very) long run, even values and norms need no

longer be assumed to be exogenously given. Their change can also be modelled as influenced by formal institutions. But in many cases, this will only work incrementally and not in one giant step (Voigt 1999b provides a very simple model along these lines).

Summing up, it has to be admitted that our knowledge concerning the growth effects of conflicting property rights systems is very limited. We have identified a number of cost components but also some possible benefits of such systems. However, carrying out a cost–benefit analysis hardly seems possible, given the difficulties of giving operational meaning to the concept of efficiency. On the other hand, it seems clear that no single system can claim to be universally efficient, as efficiency demands that some constraints are explicitly taken into account. Among them, values and norms seem to play an important role. If they differ between groups, the property rights systems deemed to be efficient would most likely also differ.

Does a country's legal history determine its future prospects for growth and development?

A simple hypothesis with regard to the possibility of stimulating economic growth via setting adequate property rights systems states that: the legal past of a country determines its economic present; in other words, little can be done to stimulate economic growth. Of late, this hypothesis has been promoted by a number of prominent authors. Note that it is not identical with the one spelled out in the last section, namely that the formal property rights system of a country should be compatible with the internal institutions of a society. The hypothesis that legal history determines a country's growth opportunities will be presented below. Some flaws in the underlying reasoning will be identified and some conclusions drawn.

Originally, La Porta et al. (1997, 1998) were interested in one particular area of property rights, namely corporate finance. They grouped countries according to their legal origin. The first distinction was between common and civil law. Within the civil law tradition, they further distinguished between the French, German and Scandinavian law systems. They found that French civil law countries have both the weakest protection of investors and the least-developed capital markets, especially when compared with common law countries.[11] Later on (La Porta et al. 1999), legal origins are taken as proxies for the political orientations of government, for the expected degree of intervention into the market, for economic freedom and for government efficiency. After adding socialist legal systems, each country surveyed is listed under one of these five traditions.

Countries are placed into one of the groups according to the origin of their company law or their commercial code. This is, of course, a sweeping generalization if one takes into account that many countries chose a kind of

super-market approach, choosing bits and pieces from a variety of legal systems.[12] Nevertheless, their results are remarkable (ibid., p. 261):

> Compared to common law countries, French origin countries are sharply more interventionist (have higher tax rates, less secure property rights and worse regulation). They also have less efficient governments, as measured by bureaucratic delays and tax compliance, though not the corruption score. French origin countries pay relatively higher wages to bureaucrats than common law countries do, though this does not buy them greater government efficiency. French origin countries fall behind common law countries in public good provision: they have higher infant mortality, lower school attainment, higher illiteracy rates, and lower infrastructure quality. … As predicted by the political theory then, the state-building intent incorporated into the design of the French legal system translates, many decades later, into significantly more interventionist and less efficient government, less political freedom, and evidently less provision of basic public goods.

Countries coded as 'German origin' are closer to common law countries, whereas countries with Scandinavian origin turn out to be quite interventionist. The underlying hypothesis seems to be: 'your legal origin is your destiny!'. As this theory has received much attention lately, we propose to deal with it in a little more detail.

First of all, La Porta et al. must assume that legal transplants are possible, no matter whether property rights systems are transplanted by conquest, colonization or voluntary adoption. If this were not the case, it would be very unlikely that attempts to transplant law would have effects decades or even centuries later. Montesquieu ([1748] 1989) famously argued that legal transplants would only be possible under very exceptional circumstances. More recently, Kahn-Freund (1974) reiterated that view with some modifications, his implicit hypothesis being that private law is more easily transferable than public law. Among comparativists, there is little consent concerning the general possibility of successful transplants. Berman, for example ([1983] 1991, p. 115) maintains that Japan and China remained largely unaffected by Western law influences; Watson (1976, p. 83) claims the exact opposite.[13]

The possibility that property rights systems can be transplanted successfully is not only crucial for the theory advanced by La Porta et al., but it also plays a crucial role for our topic: if there is a positive probability that property rights systems are not successfully transplanted, this might lead to the emergence of the coexistence of conflicting property rights, most likely the traditional system and the one intended to be transplanted.

In their more general paper, La Porta et al. basically distinguish between three law families, namely socialist, common and civil law. That socialist law tends to be correlated with an interventionist and grossly inefficient state is not counterintuitive. What is more interesting is the difference between common and civil law families. The debate is, of course, not particularly novel. In

the 1960s, Hayek (1960) and Leoni (1991) argued that common law would be superior to civil law because the latter can be radically changed overnight, whereas the former would be more stable. This would, in turn, allow private actors to form long-term expectations and to act on them, that is make long-term investments which should, *ceteris paribus*, generate higher growth rates in common law countries.[14] In the 1970s, it was then argued (Priest 1977; Rubin 1977) that common law was more efficient because parties that could deal with resources more efficiently would resort to the courts until precedent was changed to the more efficient allocation of resources.

Legal scholars seem to agree that the difference between the two families is often largely exaggerated. Posner (2002, p. 38), for example, describes the two traditions as 'convergent' (for a similar evaluation, see, for example, Zweigert and Kötz 1996). Furthermore, it was pointed out that even the most committed revolutionaries have no choice but to rely on heavy chunks of traditional law that they have set out to abolish (Böhm 1966 with regard to the French Revolution). Yet, although factual differences between the two families seem to be disappearing, the regressions by La Porta et al. are often very significant.

La Porta et al. obtain their results based on one particular area of private law, namely, company law or the commercial code, yet their results seem to carry over to the quality of government, which is based on public law. Their implicit hypothesis must thus be that private law is clearly more important than public law, constitutional law included. Differences in government performance are explained by differences in private law. It has been argued (Grady and McGuire 1999) that rulers always have incentives to make the private law as efficient as possible, the reason being that an efficient private law will increase the number of transactions and will, eventually, lead to higher tax revenue. This argument appears plausible, although it completely contradicts La Porta et al.'s approach. Both approaches must implicitly assume that private law can be neatly separated from public law. It is this implicit assumption that we do not agree with.

Private law can be called 'efficient' if it enables the actors to execute transactions at low cost and if non-compensated externalities between actors are absent. As has already been spelled out above, this means that property rights systems are not confined to private law issues; there are cases in which regulation can increase efficiency. Typically, regulation is, however, not part of private law, but of public law. Furthermore, private law, in order to be implemented, must by necessity be based on public law: the police, the procuracy, the courts are all based on public law. No matter how 'efficient' the private law might be, it will only induce growth and development if it is implemented, that is if it is backed by adequate public law. One could also formulate that an efficient private law is a necessary but not a sufficient condition for growth and development.

We know that *de iure* public law is often not identical with *de facto* public law. If that is the case, then even an efficient private law is not sufficient for growth and development. It is all the more remarkable that La Porta et al. have found a significant relationship between one particular aspect of private law and government performance. But why should *de iure* private law be a perfect proxy for *de facto* private law in the first place? One possible hypothesis is that focal points that help individuals coordinate their behaviour in a way different from that prescribed by a state-run property rights systems seem to be more readily available in private law than in public law. The coexistence of conflicting systems appears thus more likely with regard to private law issues than with regard to public law ones. An alternative system pertaining to the right to vote every so many years appears far-fetched.

An alternative way to approach the question is to assume that public law is less important for common law countries than it is for civil law ones. But why should that be the case? The suggestion that legal development would be more decentralized and that this would in and of itself already be an important element of the separation of powers is nothing more than mere speculation, at least for the moment.[15] Additional questions abound: are there certain private law families that match better or worse with certain public law families?[16] Although very little is known about these issues, it seems obvious that private law families do not comprise the same countries as public law ones: just think of the huge differences between many aspects of public law between England and the United States, the two countries that are always grouped in the same private law family.

Feld and Voigt (2003) have inquired into the economic consequences of one aspect of public law, namely, the independence of the judiciary. Focusing on the highest court of a country, no matter whether it be a constitutional or a supreme court, they find that while *de iure* judicial independence does not have an impact on economic growth, *de facto* judicial independence positively influences real GDP per capita growth in a sample of 56 countries. If the legal origin variable as provided by La Porta et al. is included, that is, if a variable based on private law is added, this changes neither the judicial independence indicators nor the economic control variables. In other words: inclusion of legal origin neither improves nor decreases the explanatory power of the judicial independence variable.

Another shortcoming of La Porta et al. is that they do not sufficiently distinguish between legal origin and a country that was formerly a colony. The former British and French colonies obviously play an important role in the results. However, having been a British or French colony, and having a British or French private law system is not necessarily the same thing. It is, at least theoretically, conceivable that a country has adopted French law without

ever having been a French colony. As Persson et al. (2003, p. 13) have recently found: 'French colonial origin is associated with less corruption, counteracting the positive effect on corruption of having a French legal system'. Corruption is, of course, only one of the many aspects dealt with by La Porta et al. Yet, Persson et al. argue that it made a difference if a country only has a French legal system or if it was – in addition – a French colony. If the French exported their implementation know-how, at least one variable of interest seems to have had offsetting effects. A more fine-grained mapping (legal origin? Colonial history?) might be in order.

Furthermore, it might be the case that La Porta et al. did not sufficiently take into account the differences between the British and the French as colonial powers. Zweigert and Kötz (1996) assert that the British did not try to replace Islamic, Hindu or unwritten African law. In India, the English courts were instructed to apply Islamic or Hindu law depending on the religion of the parties in cases of inheritance, marriage, caste and so on. In Africa, judges were to apply English law only to the extent that local circumstances permitted. The British thus did not insist on the material content of their law but did export the procedures, which enabled the emergence of a variety of common laws after the colonies had become independent.

The French, in contrast, strove to 'improve' men in the colonies and lift them up to French standards of civilization (ibid.). They thus attempted to implement the material content of the *Code Civil* in all their colonies, even if there were serious conflicts between it and, say, Islamic laws. It could thus be argued that it was not French or civil law, *per se*, that was inapt to induce economic growth and development, but the attempt of the French to implement the material content of their law even against resistance whereas the English refrained from such attempts.[17] Whereas the French (tried to) export material law, the English confined themselves to the export of procedural law, which could be 'filled up' according to local custom and tradition. This would, then, indicate that the French approach was much more likely to create conflicting property rights systems than the British approach. In other words: successfully transplanting procedural law seems to be much easier than successfully transplanting material law. Although we have identified some flaws in La Porta et al.'s approach, it has nevertheless enabled us to produce a hypothesis, namely that the French attempt to implement their material law has had far-reaching effects on the quality of government that remain significant to this day.

One could also argue that it is not origin that counts but the way in which law is transplanted. This hypothesis has been promoted by Berkowitz et al. (2002). They argue that the way the law was initially transplanted and received is a more important determinant than the affiliation to a particular legal family. They claim that the origin of a legal system as identified by La

Porta et al. (1998, 1999) might be a good predictor for what we call *de iure* legality, but not for the factual implementation of the law. For *de facto* legality, the way the law was transformed according to the specific situation of a society, whether it was imported voluntarily or enforced by a colonial power and so forth would be much more important. According to their findings, countries that have developed legal orders internally or adapted transplanted legal orders to local conditions and (or) had a population that was already familiar with basic legal principles of the transplanted law have more effective legality than countries that received foreign law without any similar predispositions.

Applied to the issue of the coexistence of conflicting property rights, this means that (i) the development of one's own property rights system or (ii) the voluntary transfer reduces the probability that conflicting property rights systems would emerge.

To sum up: although their results are quite impressive, the story told by La Porta et al. has a number of shortcomings. For example, they did not sufficiently take into account the different approaches of the French and the British as colonizers: whereas the French were determined to export the material content of their civil law, the British often restricted themselves to exporting procedures, allowing the colonies to keep using their traditional property rights system to a much larger extent. More generally, the coexistence of conflicting property rights systems appears more likely with regard to private than to public law, although the distinction is much less precise than is often assumed.

Conclusion and outlook

As pointed out in the section on the economic consequences of conflicting property rights systems, there have been instances when the existence of systems was conducive to economic growth. The 'European miracle' is an oft-quoted example. But there are also instances in which the existence of conflicting property rights systems is indicative of the inefficiency of the state-run system – the story told by de Soto (1990) about Peru is often quoted. A number of conditions that help us evaluate whether we are dealing with a case of the first or the second kind were developed. Yet, our knowledge concerning these conditions is clearly insufficient.

The last section was *inter alia* used to demonstrate that private and public law could often not easily be separated, as their functioning is closely intertwined. Whereas various legal families have been identified with regard to private law, nothing similar exists with regard to public law. It would be interesting not only to identify such families but also to work on theoretically possible and empirically realized relationships between public and private law families.

During the age of the nation-state that was delineated on ethnic grounds, conflicting property rights systems have only been of limited relevance as long as the group living under its jurisdiction shared similar values and norms and was autonomous in having them reflected in their (state-run) system. In the near future, as societies become more heterogeneous and as supranational as well as international property rights systems are striven for, conflicting systems might again become very relevant.

Notes

* I would like to thank Anne van Aaken for many stimulating discussions on this and related topics. The usual caveat applies.

1. Additionally, one should bear in mind the possibility that the extent of the groups for which a set of external institutions is valid need not necessarily be congruent with those for which a set of internal institutions is valid. Two divergences are possible: (i) a set of external institutions is valid in an area encompassing various groups that structure their interactions according to a variety of internal institutions, and (ii) a valid set of internal institutions is more encompassing than the validity of a set of external institutions; it could, for example, be the case that members of one group live in two or more nation-states. An altogether different possibility of heterogeneous internal institutions comes into play if groups are not delineated by the criterion of geographic extension. The group of diamond traders is a group interacting the world over and largely drawing on their internal institutions (see, for example, Bernstein 1992).
2. It has been argued that even autocrats have incentives to erect efficient private law systems because that would enable them to receive higher tax incomes.
3. Massell (1968) provides a very detailed and illuminating description of how the Soviets tried to make the populations in their central Asian republics give up the sharia – and how their efforts were almost a complete disaster.
4. As anyone who has ever tried to bribe a policeman after having committed a minor traffic offence in a foreign country will readily admit.
5. Although partial correlation coefficients between various measures of corruption are astonishingly high. The problem is, of course, equally relevant with regard to estimates concerning the size of the informal sector.
6. This would suggest the hypothesis that corruption levels should be rather low in quasi-anarchical societies.
7. If we had a measure of conflicting property rights systems, it would be interesting to test whether their degree is positively related to the number of ethnicities living in one jurisdiction. A measure of the so-called 'ethno-linguistic fractionalization' is readily available and has proved to have explanatory power with regard to the quality of public goods provided in Africa (Easterly and Levine 1997).
8. Empirically, this is often the case. For example: until the end of the 1970s, choosing Britain as the forum for an arbitration clause was considered a legal technical error even though English is the lingua franca of international trade and many contracts were materially based on English law (Triebel and Lange 1980, p. 616). The reason for the declining relevance of London as a location for arbitration courts was the competences of British courts that could intervene into this private jurisdiction at any point of the procedure. The fear that London as a location could lose relevance was the reason why British arbitration laws were reformed.
9. Additionally, internal institutions that reflect widely held values and norms often serve as mechanisms limiting the effects of negative externalities which are a consequence of meddlesome preferences (Sen 1970). Internal institutions might ban the consumption of drugs because the knowledge of others consuming drugs could reduce the utility level of many members of a given society. If this was the reason for the emergence of particular norms, the evolution of an alternative property rights system which increases per capita

income might, nevertheless, not be connected with improvements in individual utility levels.

10. Of course, one can argue that a reduced number of (institutional) suppliers can very well be the result of functioning competition. The concept of potential competition claims that the *possibility* of entering a market would suffice to drive producer rents down. Barriers to entry can, however, be expected to be either substantial or even prohibitive.
11. Berglöf and von Thadden (1999) have argued that the analysis offered by La Porta et al. is 'incomplete and normative conclusions are often premature'. In particular, they argue that the worldview of La Porta et al. is one 'where the main corporate governance problem is an entrenched management and weak, dispersed shareholders'. Other points of critique include: (i) the distinction between civil and common law is rather superficial; (ii) La Porta et al. use biased or misleading measures of the quality of corporate law; (iii) the underlying causality is unclear; corporate finance might just as well drive corporate law – and not only the other way around as supposed by La Porta et al.; (iv) the correlation between legal origin and financing arrangements might be driven by a third, unobserved, variable.
12. It would be interesting to rerun their regressions based on the legal origin of some other part of private law.
13. The seeming contradiction could be somewhat reduced if attention is paid to the time span various authors have in mind. Zweigert and Kötz (1996, p. 65), for example, point out that the numerous laws imported from Continental Europe to Japan did not have an imprint on legal reality but observe that this might be slowly changing. Formulated as a hypothesis and more specifically with regard to our topic: the longer the period that a transplanted property rights system is backed by the state, the lower the likelihood that conflicting systems will endure.

    A very early example for a failed attempt to transplant public law was given by the former Lord Chancellor of Henry VI, Sir John Fortescue, who fled to France and then described the differences between England and the Continent in 'A Learned Commentation of the Politique Laws of England' (first published in the sixteenth century). The superiority of the English system concerning wealth, happiness and the entire rule system was so evident that it was hard to understand why the whole world did not simply try to emulate the British law system. His answer was that the institution of trial by jury – which he evaluates positively – depended on a specific economic and social structure that was present only in England (Macfarlane 1978, pp. 179ff.).

    MacArthur's Japanese constitution is often cited as an example for a successful transplant of public law. Yet, Inoue (1991) shows that the Japanese language version of the constitution is more compatible with Japanese internal institutions than the English language version and how the misunderstandings between American and Japanese participants over the translation of the American draft facilitated the acceptance of the new Japanese constitution.
14. Napoleon, on the other hand, intended to create a *Code Civil* which was to remain unchanged infinitely. If government is able to credibly commit on this, the probability for change should be correspondingly low. It would be an interesting empirical topic to quantify legislative activity in various countries and to check whether it is higher in common law or in civil law countries.
15. If the countries with a common law origin display a higher degree of separation of powers as conventionally measured, common law might not be the independent variable driving the result.
16. The problem is that comparativists have mainly been interested in comparing private law systems. I am not aware of any 'public law families'.
17. But see, for example, Mahoney (2001) who seems to argue that French law is 'bad' whereas common law is 'good'.

References

Austin, John (1832), *Lectures on Jurisprudence or the Philosophy of Positive Law*, 1977 edn, St. Clair Shores, MI: Scholarly Press.

Berglöf, E. and E.-L. von Thadden (1999), 'The changing corporate governance paradigm: implications for developing and transitional countries', paper presented at the Annual Bank Conference on Development Economics (ABCDE), World Bank, Washington, DC: 28–30 April.

Berkowitz, D., K. Pistor and J.-F. Richard (2002), 'Economic development, legality, and the transplant effect', *European Economic Review*, **47** (1), 165–95.

Berman, Harold (1983), *Law and Revolution – The Formation of the Western Legal Tradition*, 1991 edn, Cambridge, MA: Harvard University Press.

Bernstein, L. (1992), 'Opting out of the legal system: extralegal contractual relations in the diamond industry', *Journal of Legal Studies*, **21** (1), 115–57.

Böhm, F. (1966), 'Privatrechtsgesellschaft und Marktwirtschaft' [Private Law Society and Market Economy], *ORDO*, **17**, 75–151.

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

de Soto, Hernando (1990), *The Other Path – The Invisible Revolution in the Third World*, New York: Harper & Row.

Easterly, W. and R. Levine (1997), 'Africa's growth tragedy: policies and ethnic divisions', *Quarterly Journal of Economics*, **112**, 1203–50.

Ellickson, Robert (1986), 'Of Coase and cattle: dispute resolution among neighbors in Shasta County', *Stanford Law Review*, **38**, 623–87.

Ellickson, Robert (1991), *Order Without Law*, Cambridge, MA: Harvard University Press.

Ellickson, R. (1994), 'The aim of order without law', *Journal of Institutional and Theoretical Economics*, **150** (1), 97–100.

Feld, L. and S. Voigt (2003), 'Economic growth and judicial independence: cross country evidence using a new set of indicators', *European Journal of Political Economy*, **19** (3), 497–527.

Grady, M.F. and M.T. McGuire (1999), 'The nature of constitutions', *Journal of Bioeconomics*, **1** (3), 227–40.

Hayek, Friedrich A. (1960), *The Constitution of Liberty*, Chicago, IL: University of Chicago Press.

Inoue, Kyoko (1991), *MacArthur's Japanese Constitution – A Linguistic and Cultural Study of its Making*, Chicago, IL: University of Chicago Press.

Jones, Eric L. (1987), *The European Miracle – Environments, Economies, and Geopolitics in the History of Europe and Asia*, Cambridge: Cambridge University Press.

Kahn-Freund, O. (1974), 'On uses and misuses of comparative law', *The Modern Law Review*, **37** (1), 1–26.

Kennedy, Paul (1987), *The Rise and Fall of the Great Powers*, New York: Vintage.

Kiwit, D. and S. Voigt (1998), 'Grenzen des institutionellen Wettbewerbs' [Limits of Institutional Competition], *Jahrbuch für Neue Politische Ökonomie*, **17**, 313–37.

La Porta, R., F. Lopez-de-Silanes, A. Shleifer and R. Vishny (1997), 'Legal determinants of external finance', *Journal of Finance*, **52** (3), 1131–50.

La Porta, R., F. Lopez-de-Silanes, A. Shleifer and R. Vishny (1998), 'Law and finance', *Journal of Political Economy*, **106** (6), 1113–55.

La Porta, R., F. Lopez-de-Silanes, A. Shleifer and R. Vishny (1999), 'The quality of government', *Journal of Law, Economics, and Organization*, **15** (1), 222–79.

Leoni, Bruno (1991), *Freedom and the Law*, expanded 3rd edn, Foreword by Arthur Kemp, Indianapolis, IN: Liberty Fund.

Macfarlane, Alan (1978), *The Origins of English Individualism*, Oxford: Blackwell.

Mahoney, P. (2001), 'The common law and economic growth: Hayek might be right', *Journal of Legal Studies*, **30**, 503–25.

Massell, G. (1968), 'Law as an instrument of revolutionary change in a traditional milieu: the case of Soviet Central Asia', *Law and Society Review*, **2**, 179–214.

Montesquieu, Charles L. de (1748), *The Spirit of the Laws* (Cambridge Texts in the History of Political Thought, 1989), Cambridge: Cambridge University Press.
North, Douglass (1990), *Institutions, Institutional Change and Economic Performance*, Cambridge: Cambridge University Press.
Ostrom, E. (1996), 'Incentives, rules of the game, and development', in Michael Bruno (ed.), *Annual World Bank Conference on Development Economics*, Washington, DC: World Bank, pp. 207–34.
Pejovich, S. (2001), 'Introduction', in Svetozar Pejovich (ed.), *The Economics of Property Rights*, vol. I, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. xiii–xxii.
Persson, T., G. Tabellini and F. Trebbi (2003), 'Electoral rules and corruption', *Journal of the European Economic Association*, **1** (4), 958–89.
Posner, R. (2002), 'Lecture 2: Formal versus substantive economic theories of law', in R. Posner, *Law, Economics, and Democracy – Three Lectures in Greece*, Athens: Sakkoulas, pp. 29–46.
Priest, G. (1977), 'The common law process and the selection of efficient rules', *Journal of Legal Studies*, **6**, 65–82.
Rosenberg, Nathan and L.E. Birdzell (1986), *How the West Grew Rich – The Ecoomic Transformation of the Industrial World*, New York: Basic Books.
Rubin, P. (1977), 'Why is the common law efficient?', *Journal of Legal Studies*, **6**, 51–63.
Schelling, Thomas (1960), *The Strategy of Conflict*, Cambridge, MA: Harvard University Press.
Sen, A. (1970), 'The impossibility of a Paretian liberal', *Journal of Political Economy*, **78**, 152–7.
Sinn, H.-W. (1997), 'The selection principle and market failure in systems competition', *Journal of Public Economics*, **66**, 247–74.
Tiebout, Ch. (1956), 'A pure theory of local expenditures', *Journal of Political Economy*, **64**, 416–24.
Transparency International (2000), *Source Book: Confronting Corruption: The Elements of a National Integrity System*, by Jeremy Pope, Berlin: Transparency International.
Triebel, V. and D. Lange (1980), 'Reform des englischen Schiedsgerichtsrechts' [Reform of English Arbitration Court Law], *Recht der internationalen Wirtschaft*, **9**, 616–26.
Vanberg, V. (1992), 'Innovation, cultural evolution, and economic growth', in Ulrich Witt (ed.), *Explaining Process and Change – Approaches to Evolutionary Economics*, Ann Arbor, MI: Michigan University Press, pp. 105–21.
Voigt, Stefan (1999a), *Explaining Constitutional Change – A Positive Economics Approach*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Voigt, S. (1999b), 'Choosing how to choose – the narrow path between effective constitutions and wishful thinking in constitutional choice', Discussion Paper 15/99 of the Max-Planck Institute for Research into Economic Systems, Jena.
Watson, A. (1976), 'Legal transplants and law reform', *Law Quarterly Review*, **92**, 79–84.
Weber, Max (1920), *Gesammelte Aufsätze zur Religionssoziologie I* [Collected Papers on the Sociology of Religion], reprinted (1988), Tübingen: Mohr (Siebeck).
Zweigert, Konrad and Hein Kötz (1996), *Einführung in die Rechtsvergleichung* [Introduction into Comparative Law], Tübingen: Mohr (Siebeck).

8 Property rights in common and civil law

Norman Barry

Introduction

Although in the modern world the problems of the economics of property are the same whatever the legal system in which they occur, and indeed there has been some convergence across different orders in the methods for their resolution, it is still important to understand that there are two great rival legal systems for their resolution – the common law model and that of civil law. One initial, and simple, distinction should be made. In common law countries a distinction is normally made between civil and criminal law: the former referring to tort, contract, trusts, property and so on, the latter meaning robbery, murder, rape and so on. It is true that actions under the former are brought by private individuals against other private individuals and in the latter it is the state that proceeds against wrongful behaviour under public law. Wrongful action held to be damaging to society at large as well as to private persons can be criminal. And it is also the case that common law systems typically have a body of public law which is not concerned with crime but administrative matters and proceedings are initiated by public bodies. But while, of course, civil law does make a similar distinction between private and public law, it is the very important differences between the *procedures* used by the two legal regimes over common problems of tort, contract and property with which I shall be concerned.

By civil law (see Mattei 2000, ch. 1) here I mean the legal systems that developed exclusively in Europe. They derive ultimately from Roman law; the French *Code Napoléon* (Napoleonic Code) (1804), the most influential of civil law systems, was based on the *Corpus Juris Civilis* (Body of Civil Law) compiled between 528 and 534 under the Emperor Justinian. Although not always codified, civilian law typically takes the form of a clearly articulated set of rules which is not created by the judiciary but by rationalistic planners. I shall, in fact, use civil law and code law as interchangeable expressions. The major source, though by no means the only one, of the civil law is the code itself and the commentaries on it, written mainly by law professors. It commonly has a 'natural law'[1] background and its rules have a degree of permanence and universality about them. Of course, they are frequently added to, and the additional elements supersede the prior features of the code. The Napoleonic Code was very much influenced by the French Revolution and one of its achievements, for good or ill, was to sweep away all feudal

concepts, especially in relation to property. They linger on a little even today in the common law.

A good example of the way that code system deals with a problem is, indeed, provided by property. Originally the French Code tried to produce an 'airtight' (Mattei 2000, p. 147) and unchanging definition of property and the rights it embodied. It represented a zone of individual freedom immune from the state and other persons. But from 1810, when the legal system first dealt with the problem of externalities, the whole notion has been subjected to a myriad of supplementary rules and regulations: there is a 'staggering number of detailed provisions attempting to physically separate rights of ownership' (ibid., p. 155), and a whole range of public law rules have been superimposed on the private law foundation. The Italian Code, the *Codice Civile*, was introduced a hundred years later and is not dissimilar to the *Code Napoléon*; equally important is the German Code Bürgerliches Gesetzbuch (BGB) of the nineteenth century, although it was more influenced by German legal theorists of the nineteenth century than by the French Code. Perhaps civil law has an advantage in that the original codes were not written in an age of mass democracy and therefore less likely to have been influenced by interest groups.

The common law, which is the predominant system in the United Kingdom, the United States and almost all English-speaking countries, is very different in origin and practice. Though there is always the danger of exaggerating the gap between the two systems (Mattei 1997, p. 79), the role of the judge remains significantly dissimilar (Mattei and Pardolesi 1991). Its source lies not in some master document, ostensibly designed to deal with all possible problems, but with the actual practice of the law itself. It is correctly called judge-made law in the sense that it emerges through judicial decision-making in a case-by-case manner. This might superficially be thought to give the judges a great deal of discretion. However, this is limited since they are guided almost exclusively by precedent (*stare decisis)* so that judges can, perhaps a little implausibly, be said to discover the law, not make it. One obvious problem with the common law is that it is to some extent retrospective[2] in that a person cannot know what the law is until the judge has decided a case. It does then depend somewhat on the vagaries of the judiciary rather than the absolute certainty of a code. In America, only the Supreme Court can overturn precedent and in England only the House of Lords can do this (I use England here rather than Britain because Scotland has its own legal system, which has some similarity with civil law). There is one important difference between English and American common law. In the former, almost all non-criminal cases are heard by a judge sitting without a jury. The exception is the tort of libel, which is still decided by the jury. In America, the jury decides all cases (verdict and damages), unless there is an agreement between plaintiffs

and defence. And in that country, all appeals (which normally reduce awards) are heard without a jury.

However, the role of the judiciary does make the common law much more flexible in comparison to the civil law and allows litigants to bargain their way to solutions instead of judges referring back to the code for a definitive answer to a problem. We shall see just how important this is in relation to disputes that arise over externalities. Is is also important to note that there is a unified judicial system in common law countries; there are no special administrative courts, and judges deal with cases of all types.

An important historical fact is the link between equity and the common law. Equity was originally introduced by the Chancery Court to modify a little the rigour of common law in the interests of justice but now common law and equity are treated as the same. The whole idea of the *injunction*, which is the main legal instrument for the common law in cases of externalities, emerged in equity from the Chancery Court.

Of course, the common law has not been immune from statutory intervention and judge-made law is always subordinate to an act of the sovereign legislature. Although much of English law is common law, it could be repealed overnight by an Act of Parliament and many market theorists have been distressed by the rise of statute law, especially in the twentieth century, since much of it has been disruptive of the natural, self-correcting processes of the exchange system under common law. Also, some of the common law has been codified. In America, for example, there is the Common Commercial Code, which governs much of economic life. However, even in regimes where the common law has been codified, precedent is normally acknowledged and the techniques of its reasoning retained. Furthermore, common law has been much less influenced by natural law thinking than have civil systems, although Blackstone's[3] legendary eighteenth-century statement of common law includes a panegyric on property which could have been written by a European natural law/natural rights theorist. It has been very influential on later thinking.

The United States (with the exception of Louisiana, which has the Napoleonic Code) is a common law country and many of the innovative decisions and theorising about externalities and property have come from that legal regime. However, there is a crucially important difference between America and the United Kingdom – that is, the existence of a written constitution and an extensive use of judicial review. The US Constitution is a kind of code superimposed over the common law system. It is always possible to challenge a statute from either the Congress or a state legislature for its constitutionality while in the United Kingdom, with the exception of European law,[4] Parliament is sovereign and the judiciary cannot reject a law passed by Westminster (although they may, and do, closely examine the

actions of the executive under a statute). We shall see how important a written constitution is in relation to property: the American document specifically protects property against government action in the Fifth and Fourteenth Amendments.[5] In France, the constitutional court, the *Conseil d'État*, does not provide the kind of judicial protection for property (although it is formally protected by the code) that the US Supreme Court does. It can only review statutes for their legitimacy *ex ante* but once they have passed this stage they are immune from the kind of judicial challenges that take place frequently in America, though it is said that the lower administrative courts provide some protection for property.

Overall, the common law is often thought to be more favourable to the market economy than is the civil law system, indeed the rise of judge-made contract law is closely associated with the rise of the market. It is, furthermore, not at all surprising that the law and economics movement, which has made such a great contribution to the solution to the problem of externalities via the notion of property, should have emerged in common law systems, especially that of America. Although the differences with civil law can be exaggerated, the common law system is much more pragmatic and flexible. A judge making a decision about an externality is likely to be influenced by efficiency considerations, and where there is a dispute over a property right is likely to allocate it to the person who can make the best use of it. Under civil law, however, there is likely to be an almost metaphysical dispute over who actually has the property in an absolute, ownership sense. In civil law, there is an attempt to distinguish *ownership* from *possession*, which is not a crucial feature of common law. Hence the obsession with determining the physical boundaries of property and the constant need to update a code in the light of changing circumstances. In a civil law system the judge has to refer back to the code, and its innumerable statutory modifications, and is not free to invent a new rule. This gives the impression that reasoning in such a legal order is deductive and formalistic.

But under judge-made law the judiciary is implicitly making a policy decision in an individual case (for a good example, see Cooter and Ulen 2000, p. 208). Of course, we have already mentioned a crucially important restraint on the judiciary in the common law – that is, precedent – but there is normally opportunity for creativity. But we are not concerned here with policy in the sense of political policy as determined by political parties but the kind of policy designed to make the legal and economic system work efficiently and with some predictability. Sometimes it might involve a dispute in a particular case between efficiency and indefeasible rights but, on the whole, the common law courts have been more concerned with servicing an ongoing system than with solving metaphysical issues about ownership.

Remedies

Even more important is the fact that the common law has developed remedies which are appropriate for economic problems. The most important of these is the *injunction*, which is an order that a judge gives to a litigant to refrain from an action or to perform an action. This gives the parties the opportunity to negotiate a settlement; thus if the injunction effectively allocates the property right (for example, the right to pollute) to A rather than B, then B can always negotiate and 'buy' the right through an enforceable contract. The two parties then share the surplus thus ensuring that a Pareto solution can be reached. There are also *interlocutory* injunctions where a temporary order is issued in advance of the dispute being later settled by a full trial. The injunction procedure is backed by the criminal sanction of contempt of court but this is rarely used. Of course, in the absence of an injunction, and in the clear case of a civil wrong, an action for damages can always be brought, but it is generally agreed that the injunction is more efficient. Civil law normally proceeds by actions for damages.

In civil law, instead of this single remedy being used, there is a plurality of corrective mechanisms, which adds unnecessary complexity to law, and, of course, increases costs. It should also be pointed out that the efficiency of the common law is greatly aided by its uniform court system; there are no specialised courts to handle property and externalities. In contrast, civil regimes have a myriad of courts (including administrative ones) and a variety of rules which are separate from ordinary legal processes.

With its pragmatic approach and emphasis on efficiency, the common law obviously has a clear utilitarian flavour to it and this contrasts with civil law. In the latter, the natural law heritage is clearly shown in its attitude to property. Property rights are normally defined, especially in the French Code, in absolutist terms. A person's property is a form of his/her individual sovereignty, deriving from the Roman law notion of *dominium*, and although this claim has been considerably attenuated by later additions to codes, the type of reasoning that it embodies has been retained in property disputes where, as we have already noted, there is an attempt to establish clear, physical boundaries between legitimate owners. People have, in principle, an indefeasible right to ownership and this is seen to be an essential aspect of their freedom. The fact that negotiation between affected parties, which is such a feature of common law, is less distinctive of civil law, though not entirely absent, explains why in the latter there is so much public law. The answer to a problem can theoretically be found in the code; but only theoretically, for some of the procedures and remedies of the common law have been adopted, almost surreptitiously, by civil law.

We can see the difference between the two legal orders when we consider the importance of the law of *nuisance* in the common law (see Cooter and

Ulen 2000, pp. 402–3). This derives from the English case of *Rylands v. Fletcher* L.R. 3 H.L. 330, 340 (1868) where Lord Codsworth ruled: 'For when one person, in managing his own affairs, causes, however innocently, damage to another, it is obviously only just that he should be the party to suffer'. This is part of the law of torts, not property. It developed entirely spontaneously through case law. An individual's 'right' to property is limited by constraints. In civil law there are complex rules which try to establish separate rights of ownership. But in the common law, what a person does with his or her property might affect adversely the interests of other owners and it is always possible to bring an action for nuisance against someone who has caused a harm.

Judicial decision making at common law is not so much governed by demonstrable claims derived from first principles of property but by the idea of *reasonableness* or acceptable behaviour. Judges historically could be relied on to make decisions which, for example, balanced the interests of, say, a developer who brought employment and prosperity to an area and the interests of existing property owners. In a famous case (*Boomer v. Atlantic Cement Co.*, 1970[6]) in which a cement plant was established in a quiet residential area in America, the plaintiffs failed to prevent its establishment. One would have thought that the doctrine (from equity) of 'unreasonable nuisance' would have led to the activity being enjoined. However, the activity could continue and the plaintiffs were rewarded with damages (compensation) which were significantly less than the value of the harm caused. But the legal innovation was implicitly justified on straight utilitarian grounds; it would have been unreasonable to enjoin such an economically valuable development. Of course, people disagree on what is reasonable and in this case some people thought the plaintiffs had been badly treated. An additional critical point in the case was that permanent, not temporary, damages were awarded. This meant that once they were paid the defendant was relieved of any further obligation. Had the damages been temporary the plaintiff could have kept coming back for more and that would have given the defendant an incentive to correct the problem. But the case illustrates the point of the common law, there are no 'right' answers and judges take a variety of considerations into account.

Again for sound economic reasons, the common law has developed the defence of 'coming to the nuisance';[7] although it is a feature of American common law, not English. A person might operate a business which generates a clear externality, it produces an obnoxious smell, for example. But if it has been in existence for some time, with no complaints, then, if someone should move to the area or build a house next door to the plant, he/she would not normally succeed in a nuisance action (see *Spur v. Del Webb*, 1972[8]). Of course, 'coming to the nuisance' is not always a defence and common law judges have some discretion. 'Coming to the nuisance' might be used by

someone to establish a monopoly and it is by no means always a successful defence. There is, therefore, a pragmatic approach which is quite unlike the relentless logic of property law in a civil system.

Signs of convergence

Despite the above-mentioned obvious differences between common and civil law it should be remembered that there is some similarity and even a convergence in their approaches to property. Common law has been clearly affected by statute and regulations, just as codes have. Indeed, one can see a parallel here. The code and the common law may be seen as two legal orders which have proved incapable, for deep constitutional reasons, of resisting the demands of statute. Both legal systems may be seen as models to adjudicate the conflicting claims of individuals engaged in market transactions. The rules in both legal orders are abstract, that is, they are designed (or have developed, as in common law) to deal with conflicts between persons not identified by class or political persuasion.

Statutes and regulations, however, are the product of political parties and pressure groups.[9] Sometimes public law might be necessary to solve a public bad or externality problem and at other times it might be necessary to plan directly an area and to forbid certain sorts of development. Thus we have zoning law, building regulations, 'takings' laws (or eminent domain) and outright nationalisation. All this is true of both systems; 'law', a body of stable, predictable rules, is being replaced by legislation, directives geared towards public ends.

Equally important are the similarities that are apparent between common and civil law in more technical matters (Mattei 2000, ch. 7). As we shall see, code systems have developed workable substitutes for injunctions, product liability law (through case law) now exists in Germany, and France has an emerging nuisance law. More importantly, there is considerable judge-made law in civil systems, especially in the lower courts, where there are opportunities for individuals to settle disputes by negotiation. No doubt this convergence has occurred because of the common problems that face all legal orders – especially externalities and the 'tragedy of the commons'. They are dealt with in significantly different ways and it is the task of normative legal theory to consider the varying solutions they offer in a more prosaic way than moralism or old-fashioned natural law.

The idea of property

The first important point to note is that property law in all societies preceded the state. As long as humankind has interacted economically some rules have developed governing entitlement. Demsetz (1967) describes in stimulating detail how Indian tribes developed a system of property rights when the rise

of hunting for animals threatened the viability of the stock. Open access would have been inefficient. In many cases the state simply confirmed and formalised what had developed spontaneously. This is most noticeable in the common law, even though the matter has, to an extent, succumbed to statute. Indeed, modern discussions about property largely turn on the question of the appropriateness of private or public rules for its governance. Efficiency here is crucial and there is little doubt that a system of private ownership meets its requirements better than an exclusively public one. After the collapse of communism, the former collectivist regimes quickly established (or re-established) private property (Pejovich 2001).

Private property is essential to a market order and any efficient economy has free exchange of property titles as its major feature. This means that there is an intimate relationship between contract and property; property is acquired through exchange. There is, in any legal system, a difference between movable (conventional goods) and immovable property (normally land). It is the latter that have proved to be most interesting and there is not a great deal of divergence in legal systems about the former, although I shall allude here to one or two. For example, common law does not make a great deal of difference between possession and ownership; they are treated as much the same. But in civil law, ownership is very important and a code's emphasis on it has prevented efficient property rights developing. And this has, of course, led to a centralised regulation of property. Even in England, private property was partially codified with the Property Act (1925). The economics and law movement has seriously doubted whether this codification is necessary. While private property is normally thought to be held by individuals, this is perhaps more pronounced in the common law system. Indeed, certain civil law countries (Germany and Italy, for example) have written into their codes redistributive goals of property law and write of the 'social function' of ownership (though in principle, the Napoleonic Code is not redistributive). There is nothing like this in Anglo-American law. The only interesting things here are the interactions between individual property holders.

There are some interesting differences in the two legal orders as to what counts as property and what is ownership. For example, it might be obvious that someone owns a piece of land but does he/she own all that lies underneath it? After all, there might be valuable mineral wealth (or oil) lying underground and is the landowner entitled to dig down to claim it? This is sometimes known as 'fugitive property' (see Mattei 2000, ch. 4). Only in American common law is there an unlimited right to search for it and it is feasible for the original owner to claim, for example, the oil that might seep into a neighbour's property. In civil codes, there are normally strict limitations on the rights that might be claimed here. There may be good efficiency reasons for the American rule, indeed its generosity encouraged the opening

up of the West in the nineteenth century, but there is always the danger of monopoly here: people might have an incentive to occupy land, not to use it effectively, but to hope that it might yield something of value in the future.

A further interesting difference is how the respective legal orders deal with the question of disputed ownership. If someone buys a piece of jewellery in good faith but it turns out to be stolen, who is the owner? Does the innocent purchaser have a legitimate claim? What is the best efficiency rule here? In civil law, the good faith buyer becomes the legitimate owner and the onus is on the original owner to recover it from the thief. This might, of course, be difficult. In common law systems, however, the innocent buyer has no real claim (Cooter and Ulen 2000, pp. 141–2) and the original owner retains the entitlement. In the first case, the incentive is on the owner to take good care of his/her property and in the second there is an incentive for the buyer to be very prudent about his/her purchases. In advance of substantial empirical evidence it is impossible to say which is the more efficient rule.

Adverse possession (known in civil law as *usucapio)* is recognised in all legal systems (Mattei 2000, pp. 114–17) including those civil orders that stress the absolute ownership features of property. Adverse possession occurs when somebody who has regularly made use of, for example, a piece of land owned by someone else acquires a legal claim to it. We are assuming here that the owner has not objected to the use and the user has behaved in an open manner: it was always possible for the original owner to reclaim his/her property. After a period of time, which varies across legal systems, the user acquires the right to the property by common law or in the strict wording of a code. There are good efficiency reasons for adverse possession: it leads to a better use of resources and encourages the energetic and entrepreneurial to employ their skills in productive ways and hence contribute to economic progress. The original owner is theoretically made no worse off because he/she could have stopped it at any time. There might even be doubt about the original ownership.

Indeed, the complaint is often made that capitalism does not take off in underdeveloped countries because people have no property rights: they often dwell in temporary accommodation in shanty towns, but because they have no secure property titles they have no collateral on which to secure loans that would enable them to start up businesses (see de Soto 2000). A developed adverse possession rule would have obvious utilitarian value. It would put ownership rights into the hands of the potentially productive.

The only difficulty here is determining the appropriate length of occupation before adverse possession can be legal. The normal time, in common law and code regimes, is 20 years. In Germany it is 30, but this might be because the country has a very effective land registry system so that it is always possible to know who exactly owns what.

The problem of externalities

The most pressing problem of modern market economies has been the adverse effect on third parties that economic action often produces. If a transaction between two or more people, or action by one person, produces something like pollution then others are affected and the costs are not paid for by the agents who originally caused it. There is, in effect, a redistribution of income from the victims to the perpetrator. There is therefore behind all policy the aim of internalising that cost. It should be paid for by the right person. It is here that there is a clear difference of approach between civil and common law. To repeat, the common law has always tried to do this through the law of nuisance, tort and property rights, and some of the major advances in law and economics emanate from this approach. Civil law, however, historically has been associated with administrative correction. The latter remedy is exemplified in *Décret* 15, 1810, of the French Code where the first major dent in the absolute property rights of Article 544 was made: it 'transferred to administrative agencies all collective interests threatened by industrial developments'. The law of nuisance, though not unknown now in civil law, is much less significant. Civil law starts out with the completed separation of private and public law. There has, therefore, been a vast number of detailed amendments to absolute property rights originally granted in codes. True to its individualistic foundations, the common law has tried to allocate responsibilities to private agents. In contrast, civil law tries to maintain the physical separability of private property holdings, so that a person whose property has been 'invaded' can sue for damages.

It is not, of course, possible that all externalities could be settled by either purely public or purely private law but they do represent different and rival conceptual approaches. The civil law does have an equivalent of nuisance (they are 'emissions' that do not involve invasion of physical boundaries although they clearly refer to incompatible uses) but has proved to be less creative in dealing with the problem because it is overly concerned with the physical separation of property. Aggrieved parties have to prove conclusively that there has been a clear violation of their property rights. Thus the remedies for incompatible uses under the two systems are rather different. Under common law, the injunction, along with the judicial notion of reasonableness, has proved to be very efficient while under civil law the action normally has to be based on proprietary considerations.

Under common law there is little or no attempt at physical separation of properties. If there has been any convergence of common and civil in this area it has undoubtedly been by the adoption of common law techniques by civil systems. In the German and the Italian codes there has been some admission of reasonableness and not all emissions are subject to an outright prohibition (as the separability doctrine would imply). As Mattei (2000,

p. 158) says: 'Assuming that these rights are free from obligation if not regulated by civil law is a mistake'. In other words, there is some redress in civil law for an aggrieved party in an incompatible use question without it involving a purely proprietary action. But the favoured model here by economists is definitely common law.

It was the work of Ronald Coase (1960) which showed with great rigour how solutions to externality problems could be achieved through the common law, and, implicitly the theoretical advantages that that system had over civil law. But to gain a full understanding of Coase's revolution it is helpful to look at the way externalities were handled prior to his work. The problem can be looked at from a philosophical (briefly) and an economic point of view. If there were a dispute about incompatible uses it is natural to look at the rights involved: who has the right to do what? This is traditionally the civil law approach. It leads to endless disputes about rights and, sometimes, morality – and, more practically, to constant redesigning of the code. The other, and more plausible, contribution to the externality problem came from economics (Pigou 1932). On the assumption that the external effect could not be internalised by normal market methods, it recommended government action, by taxation, regulation or outright prohibition, to correct market failure. If it is by tax, the levy should exactly equal the cost of the pollution. But this approach licenses endless interference by government.

It was Coase's achievement to show that, from an efficiency perspective, it did not matter what the property rights were or to whom they were allocated (an equilibrium could still be achieved through negotiation), and that normal, neoclassical economics, not endless philosophical argument about rights, held the key to the solution of externalities.

In a famous example, Coase considers the case of the railway company using wood- and coal-burning locomotives: the trains emit sparks that fall on a stretch of ground where farmers have planted their crops. Superficially, it might be thought that this is a question of rights; notably that the damage should be paid by the railway company which is creating the nuisance. But really it is a question of whether the company should bear the cost of not emitting the sparks (by installing some costly spark-reducing device) or whether the farmers should not plant their crops just there and suffer a cost through having to plant elsewhere. But Coase is able to show that it does not matter who has the property right since each has an incentive to negotiate an agreement. If there were no law at all, then bargaining could take place naturally. But if the case were to go to court then the winner of the action would succeed in getting the judiciary to issue an injunction, closing down the activity, from which the parties can then bargain. In a market, people will exchange until each entitlement is held by the party that values it the most highly. If the farmer were to win the case then the rail company would be

compelled, by an injunction, to stop emitting the sparks, or negotiate with the farmers. Here a liability rule would be invoked; in effect, the railway company is liable for the tort of nuisance. The injunction has clear advantages of the alternative, a propriety action which would lead to an award (or not) of damages, which would be the normal procedure under civil law.

The issuing of an injunction gives the parties the opportunity to negotiate a mutually satisfactory solution through enforceable contracts. The railway company could buy the right to emit sparks from the farmer and both sides would be better off. It would be an efficient solution without the need for incessant legal action about who had the proprietary right. It is to be noted here that Coase is implicitly rejecting the automatic assumption that externalities can only be corrected by an application of the law of nuisance which leads to damages. The parties can contract out of it. Though, as we shall see, some issues can only be resolved by damages.

Then Coase introduces something of immense importance: transaction costs. They are simply the costs of doing business: finding a suitable partner, legal costs, monitoring costs, enforcement costs and so on. Where large numbers are involved they can be immense. Imagine the costs of negotiating with the vast numbers of people affected by widespread pollution and where there are many perpetrators of the phenomenon. In the farmer–railroad example there might be a problem in that each farmer would have an incentive to avoid paying his/her fair share of the legal costs if an action were to be brought. There are also 'holdout' problems when the last person required to sign an agreement demands an extortionate price. There is, then, a strategic feature to all bargaining. Also, people might find themselves in a prisoner's dilemma situation. Who will take legal action since it is in the interest of all to avoid paying costs yet to benefit from the alleviation of a nuisance. It is only when transaction costs are zero that it does not matter who has the property right: each party has an incentive to move, unintentionally, the market towards equilibrium. But when they are significant then it clearly does make a difference who has the property right; especially to income distribution.

When transaction costs are high the most efficient remedy to an externality problem is an action for damages (derived from proprietary considerations). This has, historically, been the civil law solution. But when they are low (they can never be zero) a liability rule is the efficient remedy. Here the injunction under common law becomes crucial. A perpetrator will be compelled to cease his/her commission of the tort of nuisance.[10] With an injunction, of course, bargaining can then occur which improves every transactor's position so that a Pareto-efficient outcome is reached. Thus the judge is an important agent in what is, in effect, the assignment of property rights. But when transaction costs are low the market is itself allocating legal entitlements and we can speak of an efficient market in legal entitlements. Any misallocation by the

law can be corrected by future exchange. In the absence of bargaining, however, a fully informed judge will assign the right to the person for whom it has the most value. This can never be accurately known, although some writers have suggested that precedent is a help. Perhaps the most important role for law is to define rights clearly and to lubricate the process of free exchange. This is actually more important than allocating rights efficiently.

The Coase theorem has always been criticised as being false, the impediments to exchange will always be too great to effect a solution without coercive (state or compulsory legal) action, or it is a tautology, that is, if transaction costs are zero, a bargaining solution can necessarily be reached. Empirical research can answer the first question. As to the charge of tautology, it can be observed that some tautologies are useful. If they are logical inferences from established propositions about human nature they do direct attention to important social issues. By establishing the crucial role of common law in the resolution of economic conundrums Coase invited observers to engage in comparative legal analysis. And isn't microeconomics a set of elaborate and highly instructive tautologies?

As we have noted, all legal systems have remedies for actions that harm the property rights of a person but the common law has the advantage of a single remedy, the injunction against the perpetrator of an externality: the liability rule is invoked where it is efficient or else damages have to be sought. Indeed, the origin of the liability rule lies in the felt inadequacy and observed inefficiency of the damages action. In civil systems there is a variety of remedies and although they proclaim the absolute right of private property, in practice it is less well-protected than in the common law. This is not just because the absoluteness of property has been attenuated by endless modifications to the code but by the very plurality of the remedies for harms and, importantly, by the relative, though not complete, absence of injunctive relief. Also important is the fact that judges are constrained by the words of the code or statute and do not display the creativity of common law judges who are guided, primarily, by precedent. Overall, the judge in civil law systems does not show the implicit sensitivity to policy issues, which is such a feature of his/her common law counterpart's activities. Thus although common law systems do not normally have the formal, grandiose protections for property that civil law has, in practice, through case law, it has been more effective.

But there is something of a surrogate for injunctive relief in civil law orders: that is, subrogation (see Mattei 2000, pp. 179–80). This remedy is contained in Articles 1143 and 1144 of the French Code and Articles 2931 and 2933 of the Italian one. Basically, the court grants a subrogation when it compels a person to do something (for example, remove the wall he/she has built on the edge of another's property) or to refrain from doing something

(for example, emitting an obnoxious smell from his/her kitchen). But once the order is granted the public machinery of justice steps in and enforces the subrogation. There is little opportunity for Coase-type bargaining which can occur under common law. Under civil law, Roman law-type obligations are created which are binding.

Also, it is said to advantage the defendant. Under Article 1142 of the French Code, unperformed obligations can be dealt with by the award of damages to the aggrieved party. But what if the perpetrator pays and repeats the action, which would be technically efficient, but it is possibly more socially costly than the injunction? Under common law, the injunction forbids the harmful action: that is its utilitarian, forward-looking feature. It should not happen again and others will be deterred. It is also said that under civil law, the guilty party who repeats his/her wrongful action might eventually acquire a servitude (the right to do something on someone else's property). There are some solutions to such problems under civil law. The German Code (Sections 888 and 890) even criminalises actions in breach of a subrogation; those who disobey can be fined and even imprisoned. The French, through case law, have developed a method by which damages are permitted to rise in value for each day that the preceding wrongful situation is not corrected. It is a surrogate for the contempt power of common law and is known as *astreinte*.

The foundational point of comparison between common and civil law derives from the concern of the former with possessory remedies (Mattei 2000, p. 181). The possessor is treated as the owner but without absolute rights. Since the common law is flexible, a decision can vary so that circumstances may dictate a different verdict today from on a previous occasion. But civil law, probably from its natural law background, provides proprietary remedies: strict ownership is paramount and if that can be established, protection by the courts is guaranteed. This arises out of its concern with the separability of property. The fact that civil law is a little more tolerant of emissions is indicative of this. The resulting legal scenario then is, paradoxically, that civil law is less protective of property: certainly that of the victims of an externality. The fact that in the civil law the main remedy for a harm to property is proprietarian (damages) means that it lacks the simplicity and efficiency of common law.

The state and property

Law is not there just to protect individual private property from the depredations of other individuals but to guarantee it from the public authorities themselves. In a state subject to the rule of law, officials who use private property for social ends are supposed to be constrained by rules. Fundamental truths about human nature tell us that individuals in authority have to be subject to strict constitutions if their actions are to be consistent with the

public good. The experience of communism tells us that goodwill is not only rare, it is not enough: there has to be law. There is a case, though, in economic theory for public goods, one that is consistent with individual choice and the property rights foundations of classical liberal political economy.

But it is the view of all practitioners of this doctrine that the extension of public activity that has occurred in the past century has not been validated by economic theory or property rights legal doctrine. No legal system, whether common law or any of the varieties of civil law, has resisted the advance of public law or legislation here. The law and the courts in France provided little or no opposition to the nationalisation programme embarked on by François Mitterrand's government in the early 1980s. The European Convention of Human Rights has been equally ineffective.

It is not, however, nationalisation that has been the main target of critical lawyer–economists in recent times; in fact, there has been very little of it. But there has been a growing tendency for government regulation, and other forms of interference with private property, which, although falling short of full socialisation, amounts to a form of expropriation. I refer here to such things as 'takings laws', eminent domain, the use of the 'police power' and other forms of intervention in private property allegedly designed to advance the 'public good'. One apposite example of the milder form of expropriation is rent control which, as well as having deleterious economic effects, worsens the housing problem it was intended to alleviate, and also involves a considerable attenuation of property rights. It is a redistribution of income from owner to tenant and a denial of property. It has regularly been practised throughout the United States and Europe and property owners have not been adequately protected by the courts. The European Convention on Human Rights does ostensibly acknowledge economic rights. Article One of the First Protocol talks of the 'right to peaceful enjoyment of possessions' (a typical civil law expression) but this was of no help to the Viennese woman who brought an objection to the Austrian government's policy of rent control.[11] Also, in the United Kingdom, the Duke of Westminster claimed that a leasehold reform statute, which allowed leaseholders to buy out the freeholds of their properties at very favourable prices, seriously undermined his property rights under the Convention.[12] The European Court of Human Rights had no difficulty in both cases in finding a convenient public policy clause in the Convention which validated the government action. It was held to be 'necessary'. Courts are reluctant to interfere in government policy. Throughout the Western world economic rights have a subordinate status to civil rights, this hierarchy was formally acknowledged in a US Supreme Court decision in 1938.[13] Free speech or the right to abortion will always get better protection from the courts than property. Legislation in the former is subject to much greater scrutiny than in the latter.

The primary justification for state activity that involves the taking of property is the public good argument. There are certain goods which are wanted by individuals but which cannot be provided by the market; common defence and clean air might be the best examples. A public good is distinctive because it is non-rival in consumption unlike a purely private good (if I have the bar of chocolate, you cannot have it) and once defence (or clean air) is made available to one it is there for all; thus non-excludability means that once it is provided, non-payers cannot be barred from its consumption. These factors mean that the market is ineffective so that state provision is a Pareto improvement. Of course, many (if not most) public activities do not have these features but they are still provided publicly. In the United Kingdom, nobody is charged for major treatment under the National Health Service (though there are prescription charges and fees for dental and opticians' treatment subject to income) but obviously medical care is not a public good even though it is provided at zero price. Because there is a danger that government may expropriate private property under the guise of the public good argument, all modern legal systems try to protect individuals from a potentially rapacious public sector. But how effective is the protection?

It is not the alleged political protections, for example, democracy, that are important here (they are derisory) but the legal ones. In brief, governments should be subject to the rule of law: their actions should be regulated by principles known in advance and their expropriations limited strictly to genuine public improvements. There is some convergence here in all Western legal systems. The power of eminent domain allows government to take private property from individuals but this dangerous privilege should be subject to two principles: such action must be for the public good and accompanied by just compensation. Basically, the idea is that the protection from private wrongdoers that law offers individuals in the marketplace should be replicated in the public sphere. It is safe to say in advance that that protection has been offered more in theory than in practice.

There are two major problems involved in all takings cases: what is public use and what is just compensation? It should be remembered that in the modern world, a taking is not just an invasion of property, a physical possession by the state, but also, and more importantly, a reduction in owner value brought about by regulation. All legal systems (with the recent exception of America's) are reluctant to compensate in such circumstances, largely because it is felt, rightly or wrongly, that compensation for regulation would severely curtail what is claimed to be necessary public activity.

Although the threat to private property has undoubtedly come from eminent domain it is the case that there has been some minor improvement in the last ten years or so (see Barry 2000b) in the United States. It is particularly interesting because, although America is a common law country, it has a kind

of a code, its Constitution; although that is interpreted with common law techniques. Law is a combination of this 'code' plus much judicial creativity. The constitutional document does explicitly protect property in the Fifth and Fourteenth Amendments. The Fifth is especially important because it clearly says that 'private property shall not be taken into public use without just compensation'. For much of the twentieth century, this was interpreted in a way that favoured government action (see Siegan 1997). Municipalities could do almost anything under the 'police power' (itself not in the Constitution) with little scrutiny from the judiciary. The most abject position of this judicial passivity was reached in *Euclid v. Ambler Realty Co*. (1926)[14] in which a new state statute introduced zoning; it prohibited possible profitable development. The police power was used to protect the local environment. The police power has not been used to implement zoning in Houston, Texas, with no loss in satisfactory local urban standards.

After *Ambler* the courts were supine before the legislatures and individual property owners were at the mercy of the goodwill of political authorities. It is true that physical invasions were subject to severe tests, indeed in 1982 a statute that compelled homeowners to allow phone cables to pass through their property was invalidated,[15] but regulatory takings were not only upheld but also uncompensated. What was particularly important was the fact that owners who had suffered from partial takings, when only a portion of the value of a property was lost, had no chance of redress. But in *Nollan v. California Coastal Commission*[16] Mr and Mrs James Nollan obtained some relief from the courts. The Nollans wanted to demolish an old property and build a new two-storey dwelling but they could only get permission if they agreed to a public easement across their property. They objected and the Supreme Court ruled in their favour on the ground that the requirement of the easement did not advance the public goals.

Of even more significance was the famous case of *Lucas v. South Carolina Coastal Commission* (1992).[17] Mr Lucas had bought a beach property for $1 million, hoping to develop it into two holiday homes but a later regulation forbade that kind of project. Lucas won in the Supreme Court and was fully compensated. This was progress but there was still some doubt. He had suffered a total loss and his 'investment expectations' were completely frustrated (though he could still have picknicked on the beach). He had suffered something equivalent to a physical invasion and it might be different if it were a partial takings case. But at least the Court would now protect investment-backed expectations and would not allow a version of 'utility' to override fundamental economic rights.

The question as to whether partial takings would be compensated was settled in *Dolan v. City of Tigard* (1994).[18] Florence Dolan wanted to extend her business but was told that she would have to dedicate part of her land to

public use (for flood abatement). This was a partial taking; the value of Dolan's business was not wiped out but it was reduced. Dolan was successful in her claim for compensation. What was also significant was Chief Justice Rehnquist's comment (see Barry 2000b, p. 30) in the case. He said that the Takings Clause of the Fifth Amendment was as constitutionally important as the First and Fourteenth (which are about civil liberties).

The civil law has less spectacular cases but certainly faces the same problems and the behaviour of the legal and political authorities has been similar to America's pre-Lucas. This means that political factors have been more significant in public good questions than purely legal considerations. Article 545 of the French Code certainly does guarantee property and ensures just compensation when the public good requires the exercise of eminent domain. There is also a special statute ('Code de l'expropriation pour cause d'utilité publique') that provides some protection. Under French law the question of public use is decided by the administrative courts but compensation is determined by ordinary judges (Mattei 2000, p. 200). Most of the disputed cases go to the *Cour de Cassation* but there is also the possibility of appeal to the *Conseil d'État* on both issues. It is true that the lower courts have limited political discretion but arbitrariness remains.

The question of compensation is a deep, almost philosophical, issue. It cannot be an entirely subjective matter because then a person could always claim that a particular property had such sentimental value that it greatly exceeded the market value of similar properties. There is therefore in all legal systems a move towards compensation that reflects 'objective' market value. But what is that? In such matters it is normally a crude form of utilitarianism that prevails over fundamental property rights. Certainly regulatory losses are not compensated in any civil law system.

But there is an even greater threat to property rights in all legal systems. This is the increasing tendency to grant eminent domain powers to private persons (normally companies). And once again it is the United States that provides the most instructive examples. The most famous and notorious instance was *Poletown Neighborhood Council v. City of Detroit* (1981).[19] Here General Motors wanted to buy out a residential area to develop a new plant. It had, of course, brought much employment and prosperity to that part of Michigan. The trouble was that Poletown was a thriving ethnic community and the residents were reluctant to sell. General Motors was granted eminent domain powers and it was confirmed in the resulting case. Of course, there were good utilitarian reasons for the decision but one wonders why they should have been decisive. No doubt there was the possibility of a holdout problem if the matter had been left entirely to the market, but these issues are not insoluble by exchange methods and in this case the residents and General Motors could have reached an agreement in which both parties could have shared the surplus. One suspects that the com-

pany wanted to acquire the property cheaply and it had great political influence. In such circumstances, the civil law tradition, with its formal reasoning from written rules should have provided a better protection for property than the 'creative powers' of a common law judge. The trouble is that civil law regimes have not been solicitous in their protection of private property in the face of eminent domain.

It would be quite accurate to say that the biggest threat to property rights today comes from the government. But it always has done. Traditionally, common law systems have always developed in a way that ended the threats to one person's property owner's rights that have come from a rival, even without the aid of a Hobbesian lawgiver. However, when an established government dispossesses a private owner it does so with all the majesty of law and the superficial validation of the *Rechstsstaat* (rule of law state) the danger is especially insidious. That is what happens when takings take place under the guise of public use and subject to just compensation. But they have proved to be paper-thin protections. Only in America, and only recently there, has the law begun to exert itself on behalf of individuals. It has done so by a curious combination of a code, the Constitution, and with due reverence, but not blind commitment to, precedent. In an ideal world, everything would be safe when judges almost mechanically implemented agreed-on rules and did not let their political predilections influence their judicial rulings. That, I am sure would have precluded this alarming practice of granting eminent domain powers to private persons.

We are entitled to hope that judges might interpret a constitution, or a code, that firmly recognises property rights, in a manner consistent with the original aims of the property law. They would not then be the slaves of precedent but would overturn previous decisions that were antithetical to the rights of property. That is what the US Supreme Court is doing now in relation to regulatory takings. There is still a long way to go.

Notes

1. Natural law is the doctrine that positive, or the written law actually enforced in a legal system, is not valid law unless it is consistent with a universally objective morality. See Barry (2000a, ch. 2).
2. This means that something can be valid law without being promulgated in a statute or a case. In English common law, judges try not to make criminal law retrospective. Some common law theorists maintain that since there is always a correct answer to a case, the law is not therefore retrospective. See Dworkin (1977).
3. Sir William Blackstone, *Commentaries on the Laws of England* (1765).
4. See Barry (1994) for an account of the effect of European law on Britain's legal system.
5. The Fifth Amendment to the US Constitution states that 'Private property shall not be taken into public use without just compensation' and the Fourteenth says that no state shall 'deprive any person of life, liberty and property without due process of law'.
6. *Boomer v. Atlantic Cement Co., Inc.*, N.Y.S.2nd 312, 257 N.E. 2d 87 (Court of Appeals, New York, 1970).

7. Sometimes known as the doctrine of 'first come, first served'.
8. *Spur Industries v. V. Del Web Development O.*, 494 P 2d 700 (Arizona 1972).
9. Statutes and regulations are the product of political parties and can be analysed in public choice terms.
10. See Calabresi and Melamed (1972) for the economic differences between types of legal action.
11. *Mellacher v. Austria* (A/169) 12 E.H.R.R. 391, ECHR, 1990.
12. *James v. United Kingdom* (A/98) E.H.R.R. 123, R.V.R. 139, ECHR, 1986.
13. *United States v. Carolene Products*, 304 U.S. 144 (1938).
14. *Euclid v. Amber Realty Co.*, 272 U.S. 365 (1926).
15. *Loretto v. Teleprompter Manhattan*, CATV, 458 U.S. 419 (1982).
16. *Nollan v. California Coastal Commission*, 483 U.S. 825 (1987).
17. *Lucas v. South Carolina Coastal Commission*, 505 U.S. 1003 (1992).
18. *Dolan v. City of Tigard*, 512 U.S. 574 (1994).
19. *Poletown Neighborhood Council v. City of Detroit*, 304 N. W. 2d 455 (Michigan, 1981).

References

Barry, N. (1994), 'Sovereignty, the rule of recognition and constitutional stability in Britain', *Hume Papers on Public Policy*, **2**, 10–27.

Barry, Norman (2000a), *An Introduction to Modern Political Theory*, 4th edn, London: Macmillan.

Barry, N. (2000b), 'The constitutional protection of economic liberty', *Ideas on Liberty*, **50**, 26–30.

Calabresi, G. and D. Melamed (1972), 'Property rules, liability rules and inalienability: one view of the cathedral', *Harvard Law Review*, **85**, 1089–28.

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

Cooter, Robert and Thomas Ulen (2000), *Law and Economics*, 3rd edn, New York: Addison-Wesley.

de Soto, Hernando (2000), *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else*, New York: Basic Books.

Demsetz, H. (1967), 'Towards a theory of property rights', *American Economic Review*, **57**, 347–59.

Dworkin, Ronald (1977), *Taking Rights Seriously*, London: Duckworth.

Mattei, Ugo (1997), *Comparative Law and Economics*, Anne Arbor, MI: University of Michigan Press.

Mattei, Ugo (2000), *Basic Principles of Property Law: A Comparative Legal and Economic Introduction*, London: Greenwood Press.

Mattei, U. and R. Pardolesi (1991) 'Law and economics in civil law countries', *International Journal of Law and Economics*, **11**, 265–74.

Pejovich, Svetozar (ed.) (2001), *The Economics of Property Rights*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Pigou, Cecil (1932), *The Economics of Welfare*, 4th edn, London: Macmillan.

Siegan, Bernard H. (1997), *Property and Freedom*, New Brunswick: Transaction.

9 Can constitutions protect private property against governmental predation?

*Andrzej Rapaczynski**

Why is property a value?

There is considerable confusion about the values that property rights are supposed to serve. It is possible, of course, to value the institution of private property for its own sake. It might be fun, for example, to be able to exclude other people from the use and enjoyment of certain things. But this is not the kind of feeling that is likely to inspire great respect for the institution of private property and convince a nation's lawgivers to place it among the specially protected values in the country's constitution. The reason why private property is considered important usually refers to some other values that it is supposed to serve. And it may be useful to look at some of these values briefly, for they are quite a motley of different things and, consequently, private property may be viewed by different people as good or bad for a whole variety of different reasons.

Personality or 'expressive' theory of property

According to this view, property is important because it plays a vital role in the definition, and perhaps realization, of human personality. This view is often cited in the context of intellectual property: an author of a book, for example, has a very special relation to his creation. The book (in the ideal sense, protected by copyright laws) is not just an object in the world; it is a reflection of the author's personality: his ideas, imagination, values and so on. Indeed, the book is, in an important sense, a part of the author – his objective embodiment, as it were – and when the author is deprived of control over the use to which the book is put and the way it is made available to others, his very self may be diminished. Similarly, a house which I have dreamed of, planned, saved for, and built to provide shelter for me and my family is not for me just an object among others, a pile of stones put together, but an extension of my body, something which 'belongs to me' in a very strong sense.

The personality theory derives from John Locke's idea that man, by mixing his labor with nature, transforms and 'appropriates' it – in the sense of making it 'properly his' or a part of his own self. The personality theory was later developed further by Georg Hegel and Karl Marx (who opposed the

institution of *private* property, but not property in general). These thinkers envisaged the process of labor as a uniquely human activity capable of 'appropriation', that is, a transformation of the objective and alien natural world into a genuinely *human* environment, a habitable milieu that is part of our 'own' world in which – as a reflection of our own creative powers – we see ourselves for the first time for what we are; indeed, a world the construction of which is coterminous with our own self-realization (Rapaczynski 1987, pp. 177–217).

Property as a bastion against tyranny

This view, often associated with Milton Friedman, sees property as an important form of *empowerment* of individuals, especially (but not exclusively) when confronted with the power of the state. To own something means to be at least partially independent from others. If I hate my job, for example, I can easily quit it if I have sufficient independent means, but it may be much more difficult if my children may have to go without food and clothing or my family's standard of living is threatened with long-term decline. Even more importantly, I may think twice about – indeed quickly abandon – any idea of opposing my government, if the state owns all the means of production in my country and can consequently decide about whether I can earn a living. Also, if I want to oppose government candidates in the next election, even if I do not have enough of my own funds to finance a campaign, it is much easier to do this effectively if there are numerous other private parties of means whom I can persuade to back me.

Property, according to this view, fosters very important *political* values. It is a prop of individual freedom and a bastion against the omnipotence of the state.

The economic theory of property

According to this view, property is necessary for a well-functioning market economy and serves the objective of increasing national prosperity. By giving certain individuals or groups of individuals far-reaching control over certain resources (let us assume, along with many traditional, but also simplistic, definitions, that the control in question is absolute), the institution of private property cuts down dramatically on the transaction costs necessary for an efficient deployment of those resources.

When things are not owned – but are held in common by a large, or even unlimited, group of individuals – there is a disincentive for anyone to invest in their improvement, for if I cannot exclude others from enjoying the fruits of my investment, my return from the investment is correspondingly lower and I am less likely to make it. Thus, for example, when anyone is allowed to use a piece of land and harvest whatever fruits it bears, I am not likely to

invest as much in planting on it as I would if I could keep the whole harvest to myself.

The only way to eliminate this disincentive is to find some enforceable arrangement which imposes on everyone a certain way in which resources will be used. This could be done, in theory, by a voluntary, unanimous agreement of all who are likely to use a given resource (the piece of land in our example). But such agreements are very expensive and often outright impossible, so they are not likely to be made. There are two ways in which societies cut down on these costs of collective agreement. A *political solution* is to substitute some other rule of decision (majority rule or dictatorship, for example) for the unanimous decision about how a resource should be used, and then to use coercion (by the government) to enforce the decision once made. A *market solution* is to assign all rights to a given resource to a particular individual and then rely on that individual's self-interest to maximize the value of the resource, which the individual may do through putting it to a particular use or by trading it for another resource that the individual desires more or can deploy to a greater effect. If through such trades (which are now easier because they involve one-on-one transactions, rather than the unwieldy unanimous agreements among large groups of people), all resources are assigned to those who desire them the most (at least as measured by their willingness to pay for them) or who can best deploy them for some further productive use, the sum of the wealth produced in that society will be maximized.

The degree to which market solutions may be superior or inferior to the political ones is a matter of considerable dispute. But it is not really debatable these days that markets allocate at least some resources better than governments and their agencies, and consequently a society in which private property is abolished altogether must be, to put it mildly, very substantially poorer than one in which it is, to some extent, allowed. (Market solutions, although of course also supported by coercive state enforcement, also arguably involve less intrusive forms of coercion than political arrangements, but this probably belongs more in the previous discussion of the expressive – individual autonomy-related – and political values fostered by private property, rather than in the economic theory of this institution.)

Do values served by property justify its strong constitutional protection? A question that may be usefully asked is to what extent the values served by property are themselves important enough or insufficiently protected by the democratic processes to be in need of additional protection through constitutional restrictions on the permissible outcomes of these democratic processes. Another is whether the relation between these values and the institution of private property is sufficiently close to warrant a constitutional protection of

private property as well, or whether the values served by private property might perhaps be sufficiently protected by other means. Further, not all property rights may be of the same importance, and if so, then which, if any, might be the appropriate candidates for constitutional protection? And finally, how effective are those constitutional protections likely to be?

Not the values emphasized by the personality theory

There is little disagreement today that the protection of the integrity of human personality and of the right of an individual to personal development are values worthy of constitutional protection. Indeed, most bills of rights in individual constitutions contain a panoply of such protections. But the relation between these values and private property is too speculative and attenuated to warrant strong constitutional protection of the latter. To begin with, the aspects of private property that are emphasized in this theory seem to apply to certain special situations, and do not inhere in all forms of property that the law in all countries protects. Indeed, it is not clear to what extent the relationship of ownership is really at the core of what the personality theory emphasizes.

Suppose for the moment that the story about an artist's work being a part of his own personality or that of my house's being an embodied extension of my self is in some sense intelligible and persuasive. It is still clear that the same story cannot be made plausible for many other forms of property which are usually included in most legal protections of ownership. By most accounts, the money I have in the bank is a strongly protected form of property. But how does it reflect the creative aspects of my personality? Money is a terribly abstract commodity, each dollar being only numerically different from another. To be sure, if I am deprived of my bank account, I may not be able to lead the same comfortable life that I do now and thus may not be able to develop my personality, but this importance of money for me does not rest on the fact that it in any way reflects my life plans, ideas, or any other special facets of my personal development. Indeed, money represents an unindividualized common denominator of human needs. As such, my right to it may be, up to a point, defended by reference to some form of right to minimum human subsistence, but the *property* right, at least in our society, is almost never so intimately linked to the concept of need. Indeed, while such concepts as progressive taxation may require higher sacrifices on the part of those who are better able to bear them, most legal protections of property extend to all owners, without regard to their need for the assets involved. And whatever the merits or demerits of such a link, the justification of the right to my bank account in terms of a right to minimum subsistence neither reflects the ideas involved in the personality theory of property nor requires such a theory for its defense. In fact, a right to minimum human subsistence seems

to be usually invoked in defense of taxing (and thus limiting the rights of) property owners, rather than in defense of property itself.

One could multiply these objections by invoking the problems posed for the personality theory by inherited property and other forms of ownership that cannot easily be assimilated to the stories of the artist or the homeowner. But perhaps more importantly, these stories also do not, at least not straightforwardly, support the protection of the artist's or the homeowner's *property* right in their creations. It is a rather direct implication of the personality theory that a commercial landlord's right to his 200 apartments is not as intimately related to his personality as the homeowner's right to the house he planned, designed, built and furnished for himself – and we have seen that this, while it may perhaps be taken by someone as a blueprint for radical reform, misses an important aspect of what passes for property in most Western societies. But by the same token, the aspect of the artist's or homeowner's relation to their creations, which the personality theory intends to protect, need not involve a property right at all; indeed, restricting the protection of personality to those situations in which ownership is involved seems too narrow. A renter's home may be just as important to him as the owner's, and a painting long sold to a museum may be more important to a painter than a whole collection still in his hands. In fact, both may warrant some form of special protection – such as that provided by the concept of privacy or the French idea of *droit moral*. And once such protection is provided, it is not clear that a further *property* right adds anything of importance.

Not the values of political freedom

The political theory of property looks significantly better as a potential justification of a constitutionalization of property protection. The value of political freedom and individual autonomy are clearly worthy of the highest protection. And the connection between these values and the right to property is not tenuous – the idea that private control of wealth provides a solid protection against overreaching by the state, in terms of both assuring a certain degree of individual independence from the state's largess and facilitating effective political expression at a time when political campaigns are very costly, is empirically quite plausible. Indeed, by providing a material basis of independence from all kinds of external constraints, property may be a very important support of individual freedom and autonomy in a very broad sense, well beyond its political dimension.

It is not implausible that some such considerations played a role in the decision of the American founding fathers to provide, in the US Constitution, one of the strongest set of constitutional protections of property rights. Indeed, in the eyes of the founding fathers, property was very closely tied to the very concept of citizenship, both as providing a buffer between the citizen and the

potentially overweening state and as constituting a material interest in the affairs of the state, giving the citizen a stake in important political decisions. But it is unlikely that many would still adhere today to the rather quaint idea that constitutional (entrenched) protection of property rights, beyond what a democratic society is likely to provide through the ordinary political process, is indeed necessary for the preservation of political freedom and individual autonomy. The evolution of American constitutional jurisprudence may be indicative of the transformation of both academic and popular (one may well say: armchair) political theory informing much of the constitutional discourse in America and many other parts of the world. Property protections, which had played an enormous role in the American constitutional scheme prior to the 1930s, have been gradually narrowed down to a few relatively unimportant provisions (the Takings Clause,[1] above all), while the broad sweep of the Contract Clause,[2] the Due Process Clause,[3] and the Money Clauses[4] has been, by and large, eliminated. Indeed, property rights have often come to be seen as a potential threat to democratic sovereignty, with the associated fears of the influence of large corporations, oligarchic privileges, and the undue, indeed corrupt, impact of money on the political process.

But at the same time, American constitutional law has not been insensitive to the need to protect the individual against a potentially overweening state. On the contrary, whereas the pre-1930s doctrine relied on property protections and gave a rather narrow scope to other, more directly political guarantees of individual freedom, the post-New Deal courts have greatly expanded the reach of the First Amendment freedoms of speech and religion, the interstitial protections of privacy and other personal freedoms (such as the right to free movement), the Fourteenth Amendment guarantees of civic equality and the right to vote, and a whole panoply of rights of persons subject to the criminal process. Simultaneously with this constitutional development, the concept of individual dignity and personal autonomy found its expression in a large area of non-constitutional, but nevertheless well entrenched, legislative enactments protecting people from extreme poverty and destitution – an area that in many other countries became constitutionalized as well. The electoral process is also being legislatively transformed to provide both limits to the role played by private money and guarantees of access for the parties challenging the government.

All in all, while it may be too much to say that property rights are no longer seen as having a significant political dimension, their role in protecting individual freedom against political tyranny is rarely viewed as sufficiently central to warrant serious constitutional concerns. Most constitutions still put some limits on the state powers of eminent domain and even ban some extreme forms of regulatory confiscation, but these are hardly seen as vital guarantees of political freedom.

Property, economic prosperity and democracy

It is interesting that although economic prosperity is paradigmatically one of the main objectives of government – indeed, one to which most modern governments devote their overwhelming attention – wealth creation is not a value that is often constitutionalized. Indeed, while most constitutions limit the powers of even the most democratically elected governments in favor of individual rights (including some property rights), and sometimes in the name of other values, such as federalism, very few impose any restrictions for what may be inferred to be economic reasons. In the United States, where such restrictions had been, arguably, imposed at the time when the Constitution was adopted, they have been mostly interpreted out of the constitutional text over the next century and a half: this has been the fate of the Money Clauses, the expansive Marshallian reading of the Contract Clause,[5] and the Lochnerian commitment to *laissez-faire*, market ordering of economic life.[6]

This is perhaps surprising. For it is difficult to argue that democracies, or governments in general, are not prone to do harm to the economic prosperity of the nation. Indeed, throughout the history of political theory there has been a rather persistent fear that democracy is particularly prone to wealth redistribution that may be harmful for the country as a whole. The wealthy, who are responsible for a disproportionate part of a country's investment, are, by definition, in the minority. Without some anti-majoritarian checks, therefore, one could expect democratic majorities to be inherently prone to redistribute too much wealth and to lower both the return on capital and the rate of investment. The less the members of the majority understand the nature of economic processes, the more they are susceptible to demagoguery and feelings of resentment, and the more harmful the extent of potential redistribution.

Even assuming that democratic majorities are not uneducated plebeians, any government (or parties in control of governmental machinery), may be viewed as prone to excessive redistribution. Moving from the early anti-democratic conceptions to a more modern liberal theory, it has long been believed that governments develop interests of their own, which they pursue to the detriment of the general interest. An enlightened government will not, to be sure, adopt policies that impoverish the country to such an extent that the rulers' own revenues (counting both taxation and other privileges) will be lower than when less rapacious policies are pursued. But even the most enlightened governments, to the extent that they are self-interested, will maximize the interests of the country only to the extent that it fosters the interest of the rulers themselves (be they a democratic majority or an unelected dictator) (Brennan and Buchanan 1980, pp. 26ff.).

A liberal political view of this kind, which is quite close to the beliefs of the American founding fathers, has been developed more recently, through the application of modern economics, into a more elaborate 'public choice

theory'. According to this theory, the sphere of political action should be viewed as a market of a special kind in which individual agents, both the rulers and the ruled, are attempting to maximize their economic and other interests. In democracies, government officials tend to maximize their salaries and other perquisites as well their chances of re-election. In doing this, they respond to various constituencies which intend to obtain from the government various benefits (rents) that they cannot obtain on the market. Depending on the structure of political institutions – their 'constitution' – the pursuit of private interests by all the parties concerned may yield very different results: in some institutional environments – when, for example, there is no transparency of how the government works – it may be easier to obtain purely private benefits than in others. In this sense, the relative extent of welfare-enhancing governmental actions versus harmful redistributions may generally vary depending on the obtaining constitutional arrangements.

If this view has any plausibility, it undercuts the idea that governments, even the most democratic ones, can be simply trusted to pursue wealth-creation policies to anything like the extent desirable from the point of view of the country as a whole. Much as the corporate structure in a modern economy involves a separation of ownership from control, which creates an unavoidable agency problem for the owners, so in a modern political society, the separation of the people in whose name power is exercised from the government that exercises it creates a *political agency problem* that marks all modern states. Democratic elections, apart from any other values they may serve, may be useful as a tool for controlling political agency problems, and thus contribute to economic prosperity, but they are, by themselves, by no means sufficient to eliminate severe conflicts between private and public interests. Indeed, the rather common misidentification of the will of elected majorities, even those elected under the most 'ideal conditions', with the 'voice of the people' is one of the more persistent – and confusing – clichés left over from the less insightful branch of the eighteenth-century European political theory. In fact, only the constitutional (with lower- as well as uppercase 'c') arrangements that limit the sovereignty of democratic majorities and define the context of democratic governance can narrow (though never completely close) the gap between private rent seeking and the quest for greater overall welfare.

What role, then, is to be played by the protection of national prosperity in the constitutional scheme defining the context of democratic governance? Prima facie the answer seems to call for a very large role for such constitutional protections. Economic prosperity is indisputably a very important purpose of government. It may be difficult to constitutionalize it directly – after all, the imposition of a duty on the government to protect national wealth creation and assure long-term economic prosperity is likely to be

purely hortatory and not make much empirical sense, certainly not enough to provide a set of even vague criteria for any kind of judicial enforcement. But the objective of economic prosperity is all-pervasive in all constitutional texts, which are attempting to lay grounds for a system of governance that is hoped to be able to assure such prosperity.

The connection between private property and economic prosperity is also rather uncontroversial nowadays, even if the link was more questionable for much of the nineteenth and twentieth centuries, when utopias of egalitarian, communist and scientifically planned economies were taken more seriously than they are today. The need for a strong state, capable of providing both an infrastructure of the market arrangements and the legal system capable of protecting it, is, of course, nearly universally recognized. The extent to which markets fail and require state intervention to assure higher levels of social welfare is a disputed question, leaving room for significant disagreements about the proper role of the government in managing the national economy. But it is not really disputed that some – rather extensive – role must be left for the decentralized forces of the market if a country is to have a dynamic and growing economy. Nor is it disputed that, for the markets to work, economic agents must have secure long-term control – property rights – over the resources they deploy and a wide-ranging freedom to contract about their deployment. Insofar as that is the case, protection of property rights is a fundamental prerequisite of a successful economic order. Finally, as we have seen, this protection cannot be fully entrusted to the political branches of the government, regardless of how democratic they may be, without risking a conflict between the redistributive (rent-seeking) demands of the various political actors and the economic interest of the community as a whole. Some constitutional limitations on the powers of the majorities to lessen the force of property rights thus seem reasonably desirable.

Lest the nature of such limitations be misunderstood, it is important to note that the constitutional restraints on the ability of governments to abridge property rights can also be seen as, in a more important sense, *empowering* those governments, and not just limiting them. For one of the most important powers of governments, especially in economic matters, comes from their ability to precommit to investors not to adopt certain tax, regulatory, and other policies that might lead to an undue lowering of the return on investment in the future and a consequent impoverishment of both the government and the country as a whole. Countries with a long history of business-friendly policies have perhaps the most valuable aid in making such promises – their reputation based on a long practice of self-restraint. But other countries, those most in need of new investment, do not have it; indeed, their past history often provides more notoriety than reputation. Governments of such countries are, therefore, most in need of other devices that, like the ropes that bound Odysseus to the mast of

his ship, will credibly allow them to claim that they will not, in the future, when such moves may be politically expedient, respond to the siren calls of expropriation. Perhaps the most fundamental question of constitutional theory is whether, and if so under what conditions, *constitutional* prohibitions are credible enough to increase significantly a government's ability to precommit effectively to a set of policies the adherence to which may become against the government's interest once the addressees of those commitments have sunk significant amounts of capital in their initial investments.

In the remainder of this chapter, I shall argue the following two propositions:

1. Contrary to a long constitutional tradition, it is not likely that the kind of private property that is relevant in the context of the goal of national prosperity can be effectively enough protected as an individual right and that its articulation as an individual right raises a number of troubling considerations.
2. To the extent that this type of property can be effectively protected through constitutional restraints, such protection can best be accomplished through a number of indirect measures constraining the government's procedural or substantive freedom of setting economic policies, rather than through the traditional explicit protections of property rights.

Direct constitutional protections of property rights

There are two types of considerations that raise doubts about direct protections of property as an individual right, similar to the right to free speech or to profess the religion of one's choice. One is that direct protection of property is difficult to articulate as a legitimate individual right. The second is that such protections have proved, by and large, of only very limited effect, and all attempts to make them more effective actually lead to serious theoretical and practical difficulties.

Can property be legitimately protected as an individual right?

It is important to keep clearly in mind what is meant by the property that needs to be protected in order to assure the proper functioning of the market and contribute to the goal of national prosperity. The answer is: investment property, that is, the right of an investor to receive, after allowing for some moderate and legitimate measure of taxation, a significant portion of the market returns on his investment.[7]

This needs to be explained a bit further. Property, as I explained, is a very high degree of exclusive control over an asset. This exclusive control means

that the owner is able to appropriate all the future streams of income arising from the use of an asset and has to pay all the cost involved in, or resulting from, its use. The benefit of this is that it provides proper incentives for the socially optimal use of an asset. But whatever the use the asset is ultimately put to, it need not in any way involve any actual work or personal effort of the owner (though the owner, unless represented by a trustee, must provide some form of consent to the way in which his property is deployed). Indeed, a strong link between ownership and the actual management of assets – present in the earlier period of economic development – is highly inefficient because owners may not have managerial talents and people with managerial talents may not have enough assets. One of the most important transitions in many advanced economies, therefore, has been the increasing separation of ownership from control over productive assets (see Berle and Means 1932), and the creation of the capital markets which allow those who own assets to make them available, in exchange for a (residual or fixed) share of the returns, to those who know how to manage them. What this development means is that the 'exclusive control' that defines property is, for all practical purposes, reduced to what Anglo-Saxon lawyers call 'beneficial ownership', that is, the right to enjoy the returns, rather than exercise control over the actual use and deployment of the assets involved. In other words, the essence of productive property in a modern industrial society is reduced to the form of capital, and it is the protection of capital that is primarily dictated by the objective of national prosperity.

It is also important to note that, having taken the form of capital, property becomes, in a sense, purely 'abstract' and detached from any personal qualities of the owner. The owner's right now needs to be protected entirely independently of whether the investor is in any way 'good' or 'deserving', much as banks must pay interest on the money deposited with them, without asking whether the depositor is a good father or goes to church every Sunday. My favorite example is that of a good-for-nothing spendthrift heir who passes his whole life terrorizing servants, partying and drinking himself into a stupor on his yacht on the Riviera. If the return on his capital is too low, he (or rather his trustee or an investment adviser) will move the capital elsewhere or, if he cannot do that, will increase his consumption and other unproductive forms of behavior. The protection of this person's property is just as important from the point of view of its contribution to the objective of national prosperity as that of his more deserving co-investors.

The protection of the rascal on the Riviera is not just an extreme case to be explained as the price we must pay for protecting the property of the people who really 'deserve' it. This is not meant to deny, of course, that some forms of property, such as an artist's work or a homeowner's home, may deserve additional or separate protection because of their link to the worth of the

owner. But the point is that investment property is protected because there is a *social* need to protect it, independently of any notion of the worth or desert of the owner. Indeed, the protection of the investor does not in any way aim at *his* person. It is rather directed toward the protection of the community as a whole.

Perhaps the same could be said about freedom of speech: we accord this right to everyone, without regard to the worth or desert of the speaker.[8] Indeed, the whole point of freedom of speech is to protect the right of those with whom we disagree, often to the point of having contempt for them. Even then, we want to protect their right to speak as much as our own. The problem with property, however, is that it is not just unrelated to any personal worth or desert; it is also very unequally distributed. Indeed, unequal distribution of property is the very essence of private property from the point of view of the objective of national prosperity. To be sure, there is some research tending to show that extremely unequal distribution of wealth may not be conducive to economic development (see Bénabou 1996; Benhabib and Rustichini 1996). It is also the case, of course, that some people make more use of their right to free speech than others. But unlike in the case of free speech, which, by and large, everyone can use as much as he wants,[9] the whole edge of the right to property is its unequal distribution. For if the right is one to the *market* returns on one's investment, then inequality is written into the very core of the right. As a matter of fact, from the point of view of the security of investment, the ability to protect (and thus to attract) the huge fortunes of the few is just as, or sometimes even more, important than the fate of the savings of the many. To indulge in the fiction that the purpose of the right is to protect everyone equally, rather than the haves' right to have more than the have-nots, is to reformulate Anatole France's dictum that the law, in its majestic equality, prohibits sleeping under the bridge to the rich and the poor alike (France [1894] 1917, p. 75).

The right to market return on investment, then, precisely because it is a right to market return, accrues to the owner as a result of something that is external to the owner – the fact that others can put the owner's money to good use. In this sense, the fact that an owner happens to be a good and otherwise deserving person is just as accidental as that the right inheres in the rascal on the Riviera; they deserve it equally only in the sense that they do not deserve it at all. Indeed, for those who might be inclined to look at the world as reflecting an essentially moral order, property must appear as an 'injustice' in the same way as it is 'unjust' (that is, undeserved) that some of us are born clever and beautiful, while others are stupid, ugly, and get struck by lightning to boot.

In light of all this, framing the defense of property as an individual right may amount to adding insult to injury. Inequality of property, at least within

certain parameters, is socially beneficial, and for the sake of this common good, we should be willing to tolerate the fact that people who may in no way deserve it, have much greater access to power and other valuable goods that wealth makes possible. However, an articulation of the defense of property in terms of individual rights inevitably slips into such fatuous claims as that wealth is a reward for industry or frugality – claims that are not just false, but also offensive. Industry is, of course, rewarded separately from property – as wage from managerial or investment skills – and should not be confused with return on investment.[10] The idea that modern fortunes are a reward for frugality, in turn, is comparable, in its moral sensitivity, to the statement that the feudal privileges of birth were a reward for the nobility of character.

The ineffectiveness of property as an individual right

The problem with articulating property as an individual right is not only that it does not fit with other individual constitutional rights – which generally empower all persons equally and relate strongly to individual worth and dignity – but also that the protection of property as an individual right is mostly ineffective. To realize the nature of the problem it is best to ask what are the main threats to security of investment around the world. It is pretty safe to venture the view that very few governments today openly advocate, or attempt, large-scale – compensated or uncompensated – nationalizations, such as were fashionable half a century ago when the ideas of socialism were much more alive and when the faith in the state's ability to do better than the market was relatively widespread. Those days being largely gone – for a while at least – capital is not seriously threatened by ideologically-driven confiscations. What threatens the security of investment property in many countries – though the threat is on a widely different scale in different places – is a set of policies and practices that leave the owner's title largely untouched, but erode the value of the underlying assets to the point of making return on investment very risky, low and often non-existent. Corruption and extortion, a system of permits, licenses and exemptions administered by an inefficient bureaucracy – these are usually the first barriers. High and often unpredictable levels of taxation; a panoply of regulations that do not remedy market failures, but favor inefficient, politically powerful competitors or protect the jobs of employees; unstable currency; budget deficits funded by inflationary monetary policy; and a myriad of other exploitative and counterproductive measures that deliver short-term benefits to the ruling class, but detract from the long-term stability and prosperity of the national economy – these are the ways in which the value and security of property is primarily eroded in modern societies.

It is, of course, a good question whether any restraints written onto a constitutional parchment can effectively change such practices and improve

the climate in which a national economy may prosper. Constitutions are, after all, pieces of paper that are given reality by the institutions that undergird them: an independent judiciary, separation of powers, respect for the legal system, a panoply of civic institutions, and other intangible factors, such as the 'social capital' embedded in long-standing practices and civic associations (see Putnam 1993). While written constitutions cannot be effective without these institutions, it is also arguable that, once such institutions exist, written constitutions do not add much to their operation: property is not secure in many countries with beautiful paper pronouncements, and it is quite secure in Great Britain, without an indigenous written constitution. But even on the assumption that constitutional restraints, while certainly not sufficient to do the job, do matter and help shape the institutional arrangements within which political power is exercised, the restraints most likely to contribute to the security of property required for a successful economy are very difficult to articulate as individual rights. Budgetary deficits and inflationary policy may be more responsible for undermining the effective security of property than outright confiscations, and defining an individual property right in such a way as to include a private remedy against, say, an unbalanced budget is not likely to be either meaningful or efficacious. Nor is it likely to be workable to include in the individual right to property a significant private remedy against excessive taxation. If the abusive government practices of this kind are to be curtailed by legal means, these means must enter into the part of the constitution dealing with institutional design, either in the form of substantive restraints (such as balanced budget provisions or rules specifying the principles of taxation) or in the form of procedural requirements (such as depoliticization of certain decisions through delegation to independent agencies, provisions for a strong central bank and so on) that may make such practices more difficult. Strengthening the individual rights provisions is not likely to have more than a marginal effect.

The meaning of constitutional anti-confiscation provisions

The core of constitutional protections of property around the world are various anti-confiscation provisions, barring governments from taking private property without compensation. These provisions are sometimes seen as important defenses against state overreaching, and developing countries are expected to include them in order to reassure foreign and domestic investors.[11] At the same time, it is quite obvious that their practical effect is rather marginal, so proposals are sometimes heard that their meaning and enforcement should be properly expanded to add 'bite' to the checks on state action they are supposed to imply. The most radical of these proposals advocate that, unless administrative costs are forbidding, the state should always 'pay its way' whenever its actions have a negative impact on private property,

whether such impact is direct and involves an explicit taking of private property for public use, or indirect and affecting the value of property through regulatory provisions (see, for instance, Michelman 1967; Fischel and Shapiro 1988, p. 269).

What are the anti-confiscation provisions supposed to accomplish? Despite their apparently intuitive meaning, they are in fact highly ambiguous. The ambiguity stems from the fact that there are two very different circumstances in which government action can have negative impact on private property. The first is when a government, perfectly legitimately and in pursuit of the public interest, acts to correct some kind of market failure, such as when it, for example, enables a road to be built through land owned by a large number of people, some of whom might refuse to sell their property in order to hold out for a larger (indeed, unfairly large) share of the social surplus to be created by the new use. The road-building example is the most uncontroversial. But legitimate government regulations may also affect the value of private property, indeed, sometimes destroy it, as when an environmentally sound prohibition on marshland development reduces to zero the value of the marshlands to a commercial developer. At a limit, a government may legitimately (though perhaps, in light of what we now know, foolishly) come to the conclusion that private ownership of productive property is, in general, not conducive to overall welfare because a planned economy might do better, and decide on a policy of nationalization. Despite being seen as primarily limitations on governmental powers, the constitutional anti-confiscation provisions are, in all of these cases, designed to make it possible for governments to proceed with their policies, even if the affected owners do not agree, as long as the governments compensate the owners by paying a 'fair' or 'just' price for their property. (What 'fair' or 'just' may mean in such a context is also not a straightforward matter (see Ackerman 1977, pp. 2–4), but we shall not enter into that here.) The provisions thus *empower* governments as much as they *limit* them: by allowing the government to force a sale, they weaken private control over assets and no longer protect it with a 'property rule' (which permits only voluntary alienation), replacing it instead with a 'liability rule' (which dispenses with the owner's consent, but requires that damages be paid) (see Calabresi and Melamed 1972). What the limitation amounts to is a ban on combining certain types of welfare-enhancing measures with a form of redistribution that allocates the whole cost of the measure to the parties whose property is needed in order to confer benefits on society as a whole. (We shall postpone for later a discussion of why one might want to impose such a limitation on the government's ability to proceed with policies which are, after all, socially beneficial.)

Quite different issues are at stake in another set of circumstances in which governmental actions may have adverse impact on private property – and

perhaps set in motion a constitutional anti-confiscation provision. This occurs when a government does not act to further a public interest, but pursues selfish interests of its own or of some favored groups for which it wants to capture the advantages which those groups cannot obtain through voluntary transactions on the market. The government may build a road that is not needed or direct it through a neighborhood where it does not belong because it hopes in this way to satisfy a well-organized coalition of government contractors and their unionized employees. Or the government may forbid the development of a marshland not because development would not be the socially optimal use of that land, but because a politically influential neighboring constituency would like to leave the land undisturbed. The government may also engage in wholesale or partial nationalization to confiscate the wealth of private investors and distribute it, in the form of patronage jobs or corrupt privatizations, to its cronies and supporters. Can constitutional anti-confiscation provisions also remedy these types of evil and provide a break on inefficient and corrupt governmental activities? Can they *disempower* the government with respect to abuses of its authority, but also *empower* it to make long-term commitments to a set of policies that are sounder for the nation as a whole?

It seems rather clear that, historically, the anti-confiscation provisions were primarily designed to deal with the first type of consideration: the need to compensate those whose property rights must be invaded for the purpose of achieving some higher social good. Indeed, the American Takings Clause – the first of such constitutional provisions – explicitly conditions the government's ability to take private property, even *with* compensation, on a prior finding that such a taking is indeed for a 'public use'. American courts have for the first 150 years enforced this quite strictly and often invalidated attempts to accomplish via takings a goal of (what was then considered) illegitimate redistribution. In other words, although governments have always been considered dangerous in the liberal political tradition from which the anti-confiscation provisions derive, these provisions were not themselves designed to deal with the problem of governmental abuse. In the American constitutional tradition – and this tradition is the primary source of using the judicial department to review the constitutionality of both administrative and legislative uses of power – the economic overreaching by the government was to be controlled by other provisions, such as the Contract Clause[12] (prohibiting the states from 'impairing the obligation of contract'), the Due Process Clause[13] (prohibiting arbitrary deprivations of 'life, liberty, or property'), and the Money Clauses[14] (reserving all monetary policy to the federal government and prohibiting to all governments the issue of paper money).

How was it that the constitutional anti-confiscation provisions, from their limited origins in the American Takings Clause, came to be seen as the

primary bulwark of a constitutional defense of private property around the world? There are probably two main reasons. The first is that, as socialism began its political ascendency from an essentially oppositional and utopian workers' movement to an ideology of an increasing number of governments, the specter of nationalizations also brought forth the potential importance of the anti-confiscation provisions. Indeed, the advent of socialist governments gave the anti-confiscation provisions a sudden ideological edge they had never had before. Compensating people whose property is needed for the accomplishment of some socially worthy objective, such as a road or a park, is a natural inclination of every government not bent on attacking the very idea of private property – indeed, it is only because there is some sort of market failure to begin with that the government (or indeed private parties) is not buying the properties involved on the market, relying on voluntary alienation by the owners. But socialist governments do not view their nationalizations in the same way. They in fact tend to believe that private ownership of the means of production is not legitimate to begin with, and thus marked with original sin, much as other social evils, such as slavery and pollution, so that eliminating them does not call for any real compensation. Moreover, the very idea of compensation, especially at market rates, tends to undo much of the ideological meaning of nationalization. The idea of nationalization is to give back to the nation what belongs to it, not change one form in which the capitalist classes hold their wealth into another. Indeed, what can the expropriated capitalists do with their monetary wealth? Consume it? Take it abroad and make it work for other countries? All of these alternatives go against the grain of the very worldview underlying most socialist policies of nationalization, which nearly always include redistribution – and not just a change of title – as one of their primary objectives. And in this context, an anti-confiscation provision, especially with a compensation clause that calls for payment of the market value of confiscated assets, becomes a real stumbling block on the way to socialism. Indeed, its very inclusion in a constitution amounts to a repudiation of radically socialist doctrines as legitimate political platforms. Partial nationalizations of some strategically important sectors of the economy are perhaps still an option. But a strong anti-confiscation provision defeats the whole purpose of socialism as a comprehensive political ideology.

As I have explained, however, socialism, especially the kind of socialism relying heavily on the idea of nationalization, planned economy and other 'old-fashioned' doctrines of the last century, is no longer a primary threat to the security of investment. There may be some countries – South Africa comes to mind – where the undoing of tremendous wealth differentials rooted in a history of past injustices, may make large-scale redistributory programs, especially in agriculture, into issues of some political vitality. But more generally, the reason why countries are impoverished in today's world is not

because they are committed to a mistaken ideology hostile to private ownership of the means of production, but because their government's honesty is seriously in doubt and the existing set of institutions does not inspire sufficient confidence that such dishonesty could be effectively controlled. And a constitutional anti-confiscation provision, while perhaps not entirely without significance, is of only very marginal help in this respect.

The second reason why the anti-confiscation provisions came to be seen as a primary constitutional defense of private property is of more recent vintage; it is a response to the collapse of the older methods of constitutional control and to the increasing sophistication with which governments accomplish their redistributive objectives. Again, the developments in America are formative here. As I noted earlier, the US Constitution was originally one of the few that contained a number of provisions attempting to control the economic powers of democratic majorities. But most of these had fallen by the wayside by the middle of the twentieth century. The Contract Clause was reduced to insignificance before the Civil War.[15] The Money Clauses, at least insofar as their limitations on federal powers were concerned, were effectively repealed by judicial interpretation in the aftermath of the Civil War.[16] And the Due Process Clause review of economic legislation was decisively repudiated in the wake of the Great Depression.[17] The main problem with these efforts at constitutional protection of the market ordering of economic relations (with its concomitant defense of market returns on investment) was the difficulty of legitimating judicial control over the economic policy decisions made by the elected branches of government. In particular, judicial determination whether a particular piece of legislation was legitimately pursuing public interest or illegitimately slipping into political favoritism of some groups over others proved to be politically explosive. In times of major political upheavals of the Great Depression, judges were often extremely skeptical about the constitutionality of deep structural reforms, while politicians saw the judges' efforts at exercising judicial control as constitutionalizing personal and class ideological prejudices. The upshot was a decisive repudiation of judicial control over economic decisions and a radical expansion of the welfare state, with its much greater regulatory penetration and an unprecedented growth of the budget.[18]

With the enormous growth of the state, in time there also arose a reaction against the earlier common assumption, most common perhaps among economists and political scientists, that the government was a benevolent uncle trying to correct market failures and maximize social interest. This more jaundiced view of the role of the state took the form of a new 'public choice' theory which analysed the behavior of politicians in terms of self-interested behavior, 'rent-seeking' constituencies, 'regulatory capture', 'fiscal illusion' and other similarly skeptical categories.

The genius of the Takings Clause and other anti-confiscation provisions is that they promise to reconcile the desire to bring back some form of constitutional control over potential economic abuses by the government with a disinclination of having judges sit as arbiters of the government's economic policies. The idea is one that parallels development in other areas of law. In torts, for example, manufacturers are sometimes no longer required to conform to a myriad of judicially or bureaucratically imposed safety standards, but are instead charged, on a no fault basis, with the full social costs of accidents resulting from the use of their products and, once this cost has been internalized, are left to decide for themselves what safety measures to apply. Similarly, environmental regulation often proceeds by imposing limits on allowable pollution levels, or making polluters purchase the right to emit certain substances, or requiring certain levels of fuel efficiency, and then leaves to the parties affected how these goals may be achieved. Why not, then, control government inefficiency by applying a similar 'market' solution to government actions by making the government internalize the costs of its activities, including its regulatory activity, and forcing it to pay – via compensation – the cost that those activities imposed on the parties affected? In other words, if the anti-confiscation procedures are tightened, and the government is made, whenever administratively plausible, to 'pay its way', the government will have to take into account the true cost of its regulatory and other activities, and will not be able to fund socially inefficient forms of regulation (that is, those whose benefits of which are less than the costs).

No country has yet given a real 'bite' to this promise. There has been some tightening of the notoriously inconsistent takings jurisprudence in America, but no legal system has taken seriously the idea that government should generally 'pay its way', not just when it takes private property in a standard exercise of its 'eminent domain' power, but also when it regulates to clarify or redistribute the various entitlements held by private owners with respect to the use of their properties. True, the American law, for example, recognizes that a regulation may 'go too far' and amount to a taking,[19] but the American courts are willing to enforce this principle only in cases in which the value of the property is essentially reduced to zero.[20]

While some economists may think that states should go further in tightening their anti-confiscation policies, and some officials of foreign ministries and international organizations may actually look at constitutional 'takings' provisions as of some importance, the hopes attached to them are likely to be disappointed: even if countries were to increase dramatically the scope of their compensation requirements (which they are not about to do), the result would involve very serious social costs, but be of little value in terms of controlling governmental abuse. The reasons for this are essentially twofold: (i) the expansion of the compensation requirement will have a very unfavorable

incentive effect on the parties the government wishes to regulate; and (ii) the hopes that making the government pay more compensation might lead to a better quality of regulation is based on a flawed view of what makes the government act and is thus unlikely to be fulfilled.

The possibility of future government action, especially regulation, is one of the facts of life, much as are actions of competitors, consumers and forces of nature. As such, it is very important that the parties subject to such action have every incentive to anticipate it. Even if some use of a property, say, emitting a certain substance into the air, is considered socially largely harmless today, we want property owners to be attuned to the possibility that, say, scientific progress may show this use to be harmful, after all, or that the urbanization of the neighboring areas may increase the hazards involved. Even bad actions of the government should be anticipated, much as should lightning strikes, arson, war and other calamities. For in all those cases, unless investors anticipate those contingencies (legislation prohibiting certain uses of property, as much as actions of competitors or wars and natural disasters), they will overinvest in the assets the value of which may diminish as a result of such contingencies. But anti-confiscation provisions, by promising full compensation for lost investment, eliminate such incentives and thus impose potentially heavy social cost, unless limited to special sets of circumstances. For example, the cases of standard eminent domain exercises, those in which private land is taken for public use, often (though not always, to be sure) involve such a special set of circumstances because they introduce a serious element of uninsurable uncertainty and may therefore induce a large amount of socially wasteful risk avoidance (see Blume and Rubinfeld 1984; Blume et al. 1984; Kaplow 1986). Paying compensation in such cases may be a proper second-best solution, because the social losses from risk aversion in those cases may be higher than those from the impairment of incentives produced by compensation. But a dramatic extension of the compensation requirement, beyond its limited traditional context, is likely to be very costly.

More troubling still, such costs are not likely to be counterbalanced by any significant gains from curtailing governmental abuse. For the belief that if government has to 'pay its way' it will either better appreciate the cost of its measures or be less likely to abuse its powers is based on a mistaken assumption that governments act primarily in response to budgetary considerations. In fact, however, governments respond to *political* costs, and these are in no way simply correlated with the burdens to the fisc. Indeed, in many cases, the reverse may be true, as when a group on which the cost of a measure falls is politically very powerful, while when the cost is shifted to the taxpayer – or to a specially chosen group of politically powerless taxpayers, or, better still, inflated away through deficit spending – the political reaction is much more likely to be more muted. In fact, all government measures have their cost –

whether it is placed where it falls or shifted elsewhere and this cost should be thought of as a tax imposed by government action. Where the political cost of this tax is going to be the highest (and thus be felt most strongly by the government) is not something that can be determined a priori. One thing is certain, however: it will not be always – or even most often – the lowest when it is left where it falls, and thus forcing government to shift it will not have any uniform effect of ameliorating the quality of the government's regulatory activity.

Indirect constitutional property protections

There is a lesson to be drawn from the last point made. The real object of 'constitutional art' – for it is unfortunately not yet a science – in terms of structuring democratic institutions is to construct a set of arrangements that align the incentives of the rulers with the interests of the ruled – and of the society as a whole. The mechanics of this is to create structures that correlate the political costs and benefits of government actions with their social costs and benefits; in other words, the object is to create a transparent transmission belt that communicates properly to the politicians and the bureaucrats the real social effect of what they do and to make that effect as determinative in their decisions as possible. This is, so to speak, the supply side of political decisions. On the demand side, political institutions should provide to the ruled a way of effectively communicating their interests to the decision makers and structure their influence in such a way that the relative advantages and disadvantages that the various groups enjoy by virtue of the cost of their collective actions are properly balanced with respect to their actual political effectiveness (so that the groups facing more serious collective action problems do not have their interests underrepresented).[21] Even such a system will not eliminate all possibility of exploitation, for a majority may try systematically to shift the cost of governmental measures onto a disfavored minority, and thus a system of individual rights may be necessary for the protection of the interests of the minority. But a system in which social costs and benefits of government measures are fairly translated into a system of political costs and benefits to the decision makers is one in which policies, laws and regulations will satisfy at least the condition of efficiency, that is, their social benefits will exceed their costs.

This may not sound like a very effective protection of property rights, but in fact it is. For whenever a government can effect a transfer of resources in such a way that the benefits of the transfer (whomever they accrue to) are higher than the costs (on whomever they are placed), we can safely infer that the government is fixing some sort of market failure and that the transfer effected by the government would probably occur voluntarily, but for the transaction costs that made the market fail. Naturally, unless the losers are

compensated, there is still a distribution effect that some may consider wrong, and this is also where an additional individual right to compensation may perhaps be justified. But we should not jump to the conclusion that the losers should indeed be compensated because of an easy assumption that their property rights were sacrificed for the public good. Indeed, the word 'property' is a conclusory term, merely asserting that an asset cannot be taken without compensation. The real question is this: who should pay the cost (the tax) involved in the measure that the government is enacting? In this context, it should not be assumed that the party on whom the tax naturally falls has a right to have it shifted onto others – be they even the beneficiaries of the measure – for the tax may very well be taking away a windfall that the person affected had no right to enjoy in the first place. Does an owner's property right to an asset really include a use that is known to be socially inefficient? Isn't the whole justification of property, at least in economic terms, that it leads to an efficient allocation of resources and doesn't it fail in this particular case? Perhaps, of course, there are other, non-economic justifications as well, but as I have argued before, at least in the case of the abstract right to a market return on investment, such justifications do not strike me as being too persuasive. And as long as we stick with the economic justifications, protection of the market return on investment does not need to include returns from activities where markets fail to the point of requiring a correction, provided that such corrections are real, and not just a cover for inefficient redistributions.

So the most effective protection of property rights may not lie in recognizing it as an individual right, and it certainly does not lie in providing the owners of assets with compensation whenever they are adversely affected by government action. Indeed, the protections envisaged by the constitutional anti-confiscation provisions should be viewed as relatively restricted in scope, related to eliminating some ill effects of uncertainty related to government actions, rather than preventing governmental abuse. The best way to do that – and to protect the security of property – is to address this issue head on through the establishment (perhaps in an entrenched, constitutional form) of various structural arrangements allowing to translate better the social costs and benefits of government actions into the political costs and benefits felt by the politicians. This is not the place to explain what such arrangements may look like; indeed, many standard constraints on democracy, such as the separation of powers, maintenance of professional, depoliticized civil service, and many other similar arrangements are designed to play this very role. But there is certainly much room for improvement here, and modern constitutional theory, shying away from economic issues, could do better by paying more attention to possible reforms. In particular, constitutional constraints on the conduct on monetary and budget policies, imposition of transparent accounting measures on government, disclosure requirements concerning the

real social costs and benefits of the government's regulatory activity, limitations on the government's ability to restrict imports, exports and currency convertibility, and other measures of this kind might very well add to the economic integrity of government's regulatory activity. Moreover, these types of measures are much more likely to contribute to the security of the property entitlements relevant for the modern economy than such provisions as the requirement of compensation for the taking of private property.

Notes

* The author thanks Michael Schuman for his valuable research assistance.

1. US Constitution, Amendment V: 'nor shall private property be taken for public use, without just compensation'.
2. US Constitution, Article I, Section 10: 'No State shall … pass any … Law impairing the Obligation of Contracts'. In the early history of the American Republic, the Contract Clause was interpreted as barring the states from many retroactive, and potentially some prospective, forms of economic regulation.
3. US Constitution, Amendment V: 'No person … shall be deprived of life, liberty, or property, without due process of law'; Amendment XIV: 'nor shall any State deprive any person of life, liberty, or property, without due process of law'. In the first part of the twentieth century the Due Process Clause was read as protecting individual freedom to contract and barring many forms of economic regulation.
4. US Constitution, Article I, Section 8: 'The Congress shall have Power … To coin Money, regulate the value thereof, and of foreign Coin, and fix the Standard of Weights and Measures'; Article I, Section 10: 'No State shall … coin Money; emit Bills of Credit; make any Thing but gold and silver Coin a Tender in Payment of Debts'. Prior to the Civil War, these clauses were interpreted to prohibit the making of paper money into legal tender by both the federal and the state governments.
5. The most famous Chief Justice of the American Supreme Court, John Marshall (1755–1835), advocated strongly a prospective application of the Contract Clause that would limit the power of the state to interfere with the private ordering of economic life. His strongest views on this subject are contained in his dissent in *Ogden v. Saunders*, 25 U.S. (12 Wheat.) 213, 332–69 (1827).
6. *Lochner v. New York*, 198 U.S. 45 (1905). The *Lochner* case came to stand as a symbol for due-process-based limitations on the state's power to regulate the economy.
7. I do not intend to argue that societies have no right to control the level to which the market is allowed to determine the rate of return on investment – they may wish to limit the resultant inequalities of wealth, and many societies do it to a lesser or greater extent. All I am supposing here is that unless the state leaves a substantial portion of the market return to the owners of capital, reliance on the market will produce serious distortions in the allocation of resources that are likely to extract a price in terms of general welfare.
8. Societies differ somewhat in how far they are willing to accord the right of free speech to people who hold views that are recognized as contemptible by most of the other members of the society. The statement in the text is most applicable to the United States which treats the right to free speech as close to absolute, including the right to profess Nazism, racism and violent overthrow of the government.
9. I abstract here from the arguments about the *effectiveness* of one's speech: some may not have access to certain media of communication that make one's voice better heard than others. But this is not, properly speaking, an objection against freedom of speech. It is just another objection to the inequality of property.
10. One can, of course, express wages as return on human capital. While the protection of those investments is also a very important aspect of a successful economy, it is not usually framed in terms of a *property* right.
11. A survey of 100 countries around the world shows that only a handful have no explicit

constitutional provision protecting property owners against expropriation. Among these are some countries, such as Austria, Canada, or Israel, in which property rights are in fact very secure. All post-communist countries in Europe and Asia have the appropriate constitutional provisions (with some ambiguity in Turkmenistan). Some of the poorest countries, such as Cameroon, Egypt and Ethiopia have full-blown anti-expropriation provisions.

12. See note 2 above.
13. See note 3 above.
14. See note 4 above.
15. The turning point came in *Ogden v. Saunders*, 25 U.S. (12 Wheat.) 213 (1827) and became confirmed in *Proprietors of Charles River Bridge v. Proprietors of the Warren Bridge and Others*, 36 U.S. (11 Pet.) 420 (1837). *Stone v. State of Mississippi*, 101 U.S. 814 (1879) put a seal on its insignificance.
16. *Knox v. Lee*, 79 U.S. (12 Wall.) 457 (1870); *Trebilcock v. Wilson*, 79 (12 Wall.) U.S. 687 (1871); *Julliard v. Greenman*, 110 U.S. 421 (1884).
17. *Nebbia v. New York*, 291 U.S. 502 (1934); *West Coast Hotel Co. v. Parrish*, 300 U.S. 379 (1937). Any hopes that anything was left of the so-called 'substantive due process' was effectively killed in *Williamson v. Lee Optical*, 348 U.S. 483 (1955).
18. The share of government in the GDP of the United States grew from 5 per cent in 1902, to 10 per cent in 1922 to nearly 35 per cent in 1994. See Steuerle (1998, p. 66). In that same year, the share of government in the GDP of other developed countries was often much higher: 53 per cent in Austria, 51 per cent in Belgium, 46 per cent in Canada, 58 per cent in Denmark, 52 per cent in France, 53 per cent in Italy, 49 per cent in Norway and 65 per cent in Sweden. The numbers are slightly lower today. See OECD *Economic Outlook* (2002), Statistical Annex Tables, Table 26 (also available at http://www.oecd.org/oecd/pages/document/displaywithoutnav/0,3376,EN-document-notheme-1-no-no-37571-0,00.html).
19. The seminal case in the United States is *Pennsylvania Coal Company v. Mahon*, 260 U.S. 393; 43 S. Ct. 158; 67 L. Ed. 322 (1922).
20. See, for instance, *Penn Central Transportation Co. v. New York City*, 438 U.S. 104, 124, 57 L. Ed. 2d 631, 98 S. Ct. 2646 (1978).
21. So much for the simplistic assumption that the most fundamental principle of a working democracy is 'one person, one vote'. See *Baker v. Carr*, 369 U.S. 186 (1962).

References

Ackerman, Bruce A. (1977), *Private Property and the Constitution*, New Haven, CT: Yale University Press.

Bénabou, R. (1996), 'Inequality and growth', in B. Bernanke and J. Rotemberg (eds), *NBER Macroeconomics Annual*, Cambridge, MA: MIT Press, pp. 11–76.

Benhabib, J. and A. Rustichini (1996), 'Social conflict and growth', *Journal of Economic Growth*, **1** (March), 125–42.

Berle, Adolf and Gardiner Means (1932), *The Modern Corporation and Private Property*, Riverwoods: Commerce Clearing House.

Blume, L. and D.L. Rubinfeld (1984), 'Compensation for takings: an economic analysis', *California Law Review*, **72** (4), 569–624.

Blume, L., D.L. Rubinfeld and P. Shapiro (1984), 'The taking of the land: when should compensation be paid?', *Quarterly Journal of Economics*, **99** (1), 71–92.

Brennan, Geoffrey and James M. Buchanan (1980), *The Power to Tax; Analytical Foundations of a Fiscal Constitution*, Cambridge: Cambridge University Press.

Calabresi, G. and A.D. Melamed (1972), 'Property rules, and inalienability: one view of the cathedral', *Harvard Law Review*, **85** (6), 1089–128.

Fischel, W.A. and P. Shapiro (1988), 'Takings, insurance, and Michelman: comments on economic interpretation of "just compensation"', *Journal of Legal Studies*, **17** (2), 269–93.

France, Anatole (1894), *The Red Lily*, English translation (1917), New York: Modern Library.

Kaplow, L. (1986), 'An economic analysis of legal transitions', *Harvard Law Review*, **99** (3), 509–617.

Michelman, F.I. (1967), 'Property, utility, and fairness: comments on the ethical foundations of "just compensation" law', *Harvard Law Review*, **80** (6), 1165–258.
Organization for Economic Cooperation and Development (OECD) (2002) *Economic Outlook*, **72** (December).
Putnam, Robert D. (1993), *Making Democracy Work*, Princeton, NJ: Princeton University Press.
Rapaczynski, Andrzej (1987), *Nature and Politics; Liberalism in the Philosophies of Hobbes, Locke, and Rousseau*, Ithaca, NY: Cornell University Press.
Steuerle, C. Eugene (1998), *The Government We Deserve*, Washington, DC: Urban Institute Press.

10 Property rights systems and the rule of law

*Ronald A. Cass**

Introduction

Tolstoy's novel, *Anna Karenina*, starts famously with the observation that 'All happy families resemble one another; every unhappy family is unhappy in its own fashion'. The opposite is more nearly true in respect to the rule of law. Though many societies with differing governance structures and legal systems adhere in their own ways to the rule of law, societies that derogate from it do so in more similar fashion. In other words, it is easier to identify departures from the rule of law than to explain why particular actions conform to it.

The rule of law matters to people around the world because it is a concomitant of a society that is successful and, in all likelihood, just (see Harvey 1961; Barnett 2001, pp. 136–44; Cass 2001, pp. xi–xii). It does not guarantee justice or social welfare, but it does correlate with justice and social welfare (under virtually any accepted definition of those terms). That is why the concept has such broad appeal.

A critical aspect of the commitment to the rule of law is the definition and protection of property rights – rights to control, use, or transfer things (broadly conceived), including rights in intangibles such as intellectual property. Societies in which it is relatively easy to secure property rights, to protect them against infringement, to gain recompense when rights are infringed, and to transfer property rights in whole or in part to individuals who value them more highly, are more likely to succeed (see, for instance, O'Driscoll et al. 2001). Of course, the substance of the rights matters. Societies that are relatively friendly to property, not only giving it security but also providing broad scope for the use of property according to its owners' desires, will also have an advantage (ibid.).

Substance aside, however, the degree to which the society is bound by law, is committed to processes that allow property rights to be secure under legal rules that will be applied predictably and not subject to the whims of particular individuals, matters. The commitment to such processes is the essence of the rule of law (see, for instance, Dicey 1915, p. 198; Oakeshott 1983, p. 119; Scalia 1989).

Although societies differ markedly in their commitment to the rule of law, the distinctions often are less clear than might at first blush appear. The ways in which systems manage changes in property rights and in legal rules that

affect property rights, along with the ways in which systems constrain official discretion, are the keys to the effectiveness of the rule of law. But the rule of law does not bar change nor does it forbid discretion. Change is a natural part of any legal system, and efforts to limit change must be seen not as ends in themselves but as part of a larger framework for assuring predictable, valid, law-based governance. Discretion to effect changes in the nature of property rights seems an inevitable part of any property rights system. And the division between systems that conform closely to the rule of law and those that depart from it will be tied less to whether discretion to shape and alter the law exists than to the nature of the discretion, to its concentration in few or many hands and its relation to other authority and to other legal and practical constraints. Conformity to the rule of law in the end cannot be measured in discrete increments but must be viewed as the product of a set of related considerations.

Property and property rights systems

The word 'property' does not strike the listener's ear as a particularly abstruse term. We think we understand it instinctively. Property is a thing. Or things are property. But, of course, when we try to be more precise, to specify whose property a thing is, and what aspects of it are proprietary, against whom, for what purposes, we realize that property is not so readily self-defining after all (see, for instance, Singer and Beermann 1993). And if we think a bit more, we realize as well that property does not consist only of things, at least not only of tangible things. We call certain types of ideas – or at least certain forms of their expression – intellectual property, and craft property rights in them. We recognize claims for money, for services that are equivalent to money, or for employment that earns money under the rubric of property rights.[1]

The line between property rights and other rights is contestable. This reflects the fact that the notion of property itself – what it is, how it should be thought of, who should have what property rights – has been a subject of controversy across centuries. Philosophers (along with economists, historians and others) have argued whether property is something natural, existing prior to government recognition of rights, or instead is a positive construct of government (see, for instance, Christman 1994; Pipes 1999, pp. 5–29). They have contested the strength and provenance of a deep-seated historic impulse to stake out individual rights in property, with contrasting claims that the natural order is not of individual but of communal rights (see, for instance, Berry 1980). And they have debated the proper basis for and scope of rights to realty and to personal property, to intangible goods and to government largess (see, for example, Donohue 1980; Pipes 1999, pp. 36–63, 225–64; Komesar 2002, pp. 125–55). Modern-day scholars continue the arguments, drawing threads

forward from hoary engagements to mix with more current assessments of the effects of inequality or of miscast incentives (see, for instance, Hayek 1960; Reich 1964; Demsetz 1967; Rawls 1971; Nozick 1974; Christman 1994; Barzel 1997; Merrill and Smith 2001; Murphy and Nagle 2002).

Although the definition of property and delineation of the line between property rights and other rights is contestable, the core notion of property focuses on things, such as land, to which rights may be given as against the world. Where other rights often rest on the existence of some duty that is tied to particular behavior toward a more limited set of individuals, property rights tend to function as rights against the world in respect of the control, use and disposition of things.[2] And often those things are situated, physically or as the result of legal rules, so that a single set of government decision makers will have influence over the value of property rights.[3] That is the type of property – and the aspect of property rights – that is most important for inquiry into their connection to the rule of law.

Property rights systems and the rule of law

This chapter tells the tale of two systems – but not truly two separate tales. Western democracies generally have adopted rules that are quite successful at limiting official discretion to alter property rights in unpredictable ways, at promoting stable and secure rights, and at limiting governmental interference with the most productive uses of property, including voluntary transfers of property to those persons who will pay most for it (at least inferentially an indication of higher value). A legal system that allows individuals to order their lives, their personal behavior, and their business conduct secure in understanding the rules that will apply to them provides a critical spur to the investments of money, of energy, of talent that promote progress in human endeavor. Far more than natural resource endowments, sound law and government are markers for a society's success – both for its material success and for its citizens' broader sense of well-being (see, for instance, O'Driscoll et al. 2001).

To be sure, there is far more regulation of economic activity – and, hence, necessarily of the use of property – in these nations than was common 50 or 75 or 100 years earlier. And the increase in regulation reasonably can be thought to reduce the value of property from what it otherwise might be.[4] Even though many forms of government activity – support for infrastructure, transportation, availability of police protection, support for a well-functioning legal system, provision of a stable currency, and so on – can raise the value of property, still other activity must be recognized as interfering with property owners' interests in order to promote other ends. Some estimates put the cost of economically burdensome regulation at staggering levels for even the least aggressively regulated advanced economies.[5]

Nonetheless, the legal imposition on property owners in these economies has been slight enough that, together with the productivity of resources and the distribution of human resources and capital, the value of property in these societies has reached lofty levels. The total value of real property (obviously only one form of property) in the United States, for example, is estimated to be between $15 trillion and $20 trillion.[6] Stock market values, which overlap somewhat with real property but also largely represent other business property, intellectual property, and other intangible value, may amount to more than $11 trillion (taking the value of shares listed on the major US exchanges).[7]

Beyond factors such as natural endowments and the choices made in crafting substantive legal provisions, the value of property in advanced economies is testament to the societies' commitment to the rule of law. Practical commitment to rule of law values is part of the economic success story, part of the value of property and of property rights. It is part of what sets Western democracies (a term that is not strictly limited to Western nations or to governments that are fully democratic) apart from many other nations.

The core concept of the rule of law includes an understanding that society should be ordered around a set of laws that apply similarly to all on the basis of principles deducible from the rules and not dependent of the identity of the rulers. It includes the expectation that rules of law that seem on the face of it to apply to all will in fact be read similarly and applied similarly to all – that what seems to be a universal rule will find similar application to the rich and the poor, to high and low born, to those from different religions and tribes and locales, different political parties, different skin colors, different families.

The base proposition for the rule of law is that, in the formulation given by John Adams and David Hume, it intends 'a government of laws, not of men'.[8] For the laws to govern, for them to play the decisive role rather than for the particular individuals in power to do so, the laws – along with the individuals who apply the law and the institutions that mediate application of the laws – must provide reasonable certainty about the meaning of the rules that govern our conduct. And the reasonable certainty about the rules must be rooted in understanding of the rules as written rather than detailed knowledge of which individuals will ultimately be applying the rules to us (see, for instance, Fuller 1969, pp. 38–81; Schauer 1991).

Hayek (1944, p. 80) framed the point in especially strong terms, saying that the rule of law 'means that government in all its actions is bound by rules fixed and announced beforehand – rules which make it possible to see with fair certainty how the authority will use its coercive powers in given circumstances and to plan one's individual affairs on the basis of this knowledge'. One can argue the contours of the Hayekian requisites for the rule of law, the degree to which rules must be fixed or the command that *all* government

actions be bound in advance by rules. But Hayek's essential point is correct – that the rule of law directs government to provide fair notice of what rules it will apply, that the rules must allow individuals to plan their affairs with a reasonable understanding of the rules that set the bounds to lawful behavior.

The classical rule-of-law conception propounded by Hayek and others contains four elemental components (see Cass 2001, pp. 3–19). These are rule fidelity (law appliers' engagement in applying law), principled predictability (foreseeability of rules' application and meaning), rule validity (the derivation of rules from valid legal authority), and external authority (acceptance of someone other than the rule applier as the source of the rule to be applied). None of these four elements is completely obvious, and each of them can be problematic in some settings.

Meeting these elements – coming within the classical definition of the rule of law – ensures that law is not the whim of an individual, that law is not administered in ways that dramatically empower government officials at the expense of private citizens, that the law's strictures are not unduly difficult to anticipate. These elements of the rule of law cohere with a system of legal governance, of law-boundedness, that allows substantial scope to individual judgment (and, thus, to varied individual values) in charting the course of ordinary affairs (Oakeshott 1983). Because these elements of the rule of law produce the sort of predictability Hayek sought, they allow individuals to adapt to legal rules in ways most likely to improve social welfare (at least as assessed by a measure of summed individual values) or, put in less positive terms, in ways most likely to minimize social cost. Further, because the rule of law thus defined militates against the sort of whimsical, biased and dictatorial impulses that often correlate with welfare-reducing rules, the rule of law can be associated with some substantive welfare-enhancing qualities (Cass 2001; O'Driscoll et al. 2001).

These rule-of-law elements do not, however, ensure that laws are wise or just. The human mind has been little able to devise precepts that provide coercive power to any group, under any structure, and ensure that it will be used exclusively in ways that are wise and just. The focus of the rule of law is not to provide such assurance – neither through procedural requisites nor through definitional assertion – but merely to ensure law-bounded qualities that tend in the direction of better and more just legal systems (see Raz 1979, pp. 213–14, 224–6; Barnett 2001, pp. 136–44; Cass 2001, pp. 15–17). Adherence to the rule of law slows down changes in the system, increases the foreseeability of change, makes change less the product of one individual's will than of the more regularized and intricate interweaving of different wills and priorities. The rule of law is a commitment to limitations that guarantee greater stability rather than any specified end-point for law and for government. This process commitment is not the entirety of what one might desire

of a legal system – the substance of the rules surely matters as well – but it is an important goal and one that needs attention even in the best, most law-abiding systems, systems that will tend to promote human liberty, security and economic opportunity.

Property rights systems and arbitrary rule

People around the world instinctively recognize the differences among nations' commitment to the rule of law. Western democracies generally rank high on this commitment and are characterized by a set of governance structures that limit individual, official power. But differences across nations, though intuitively evident, are less readily captured by description of what does and does not comport with the rule of law than might initially be thought. Both the variance and the difficulty of bold pronouncements about the qualities of a legal system that demonstrate adherence to or departure from the rule of law can be seen in comparing the situations in Zimbabwe and the United States – one system that appears strongly in derogation of, and one that appears to operate strongly in accord with, the rule of law.

Property rights in land in Zimbabwe

In 1965, the white minority government of Ian Smith declared the British colony of Rhodesia (now Zimbabwe) independent of Britain. The Smith government represented white citizens (who were a very small minority of the population, perhaps 2 or 3 per cent) and sought to maintain white supremacist policies. Policies long in effect in Rhodesia had limited the right of black African citizens to own land, except in certain designated reserves, and land ownership consequently was heavily skewed toward whites. The breakaway from Britain in 1965 sparked an armed struggle between the Smith government and two black nationalist movements, known generally in the West by the acronyms ZANU (Zimbabwe African National Union) and ZAPU (Zimbabwe African Patriotic Union). The inequality in land holdings was both a symbolic and pragmatic factor in the civil war.

The war ended in 1979 with the victory of the nationalist groups and an agreement (the Lancaster House Agreement) negotiated among the warring groups and the government of the United Kingdom, the former colonial power. The UK, which had not recognized the Smith government's claim to independence and had imposed sanctions against the regime, formally recognized Zimbabwe as an independent Commonwealth nation in 1980. ZANU and Robert Mugabe won elections to form the new government, trumpeting land reform as a prominent part of his platform. Mugabe has held office ever since.

Although land reform did occur, it proceeded at a measured pace. During the 1980s, the government acquired about 3 million hectares of land and

redistributed that land to about 50 000 families (Human Rights Watch 2002a, p. 6). Some in Mugabe's government complained that too little was being done and blamed the overhang of colonial rule (Human Rights Watch 2002b, p. 4). The Lancaster House Agreement contained a restriction on forced land redistribution, stating that for ten years (to 1990) no land would be acquired by the government of Zimbabwe unless there was a willing seller and the government promptly paid 'adequate compensation' (Human Rights Watch 2002a, p. 6). Following expiration of that stricture, Zimbabwe's constitution was amended to allow the government to take land without the owner's consent. In 1992, legislation provided a new mechanism for calculating the price to be paid for property taken by the government, one that gave substantial power to a small group of administrators.

Despite the change in law, the pace of land reform actually slowed in the 1990s, with government responsible during the decade for settlement of perhaps 40 per cent as many families on less than one-third of the land transferred in the 1980s. One reason is that the funding for the land purchases of the 1980s very largely came from the UK and other Western nations through grants that had been exhausted by the end of the 1980s (ibid., p. 7). Outside commentary – widely reported, though denied by the government of Zimbabwe – also suggests that much of the benefit of government land reform programs went to well-placed officials and influential supporters of President Mugabe rather than to the poor Zimbabweans for whose ostensible benefit the reforms were designed (Mitchell 2001, p. 599; Human Rights Watch 2002a; see also Zimbabwe government website). Poorer Zimbabweans and soldiers who had fought for ZANU during the civil war took matters into their own hands in many parts of the country, occupying land owned by white farmers without formal government sanction.

Responding to pressure from constituents, Mugabe's government proposed a reformed constitution that would, among other things, allow the government to take land without payment to the owners. The new constitution, however, was defeated in a popular referendum. Nonetheless, in the spring of 2000, the ZANU-dominated parliament adopted the amendment providing that the government could condemn and acquire land free of any obligation to compensate the owners, asserting instead (unilaterally) that Britain had the obligation to provide recompense.[9] And President Mugabe, exercising authority to enact legislation temporarily under a law giving the president extraordinary powers to deal with emergencies, amended the Land Acquisition Act in accord with the newly altered constitution to permit uncompensated takings. Six months later, the parliament adopted legislation to the same effect (Human Rights Watch 2002a, pp. 10, 13; 2002b, p. 15).

The changes in the law were accompanied by a surge in land seizures by former ZANU fighters and others. Within a year, about 100 000 squatters

occupied almost half the farms targeted by the government for acquisition (Human Rights Watch 2002a, p. 11). Within two years, according to some reports, as many as 300 000 families occupied some 11 million hectares of land formerly owned by white farmers.[10] While some white farmers abandoned their property in the face of rising violence, others resorted to armed resistance, long a part of white minority rule in Africa and particularly in Zimbabwe–Rhodesia.[11] Press reports link the violent seizures of farms to groups associated with Mugabe.[12] An umbrella group for larger commercial farmers, though lacking the political standing to reverse Mugabe's policy, both tried to persuade the government to alter its policy (attempting to negotiate an expanded, voluntary land reform program) and also went to court to block the land seizures. In December 2000, the Supreme Court of Zimbabwe agreed with the farmers that the government could not take away their property without compensation. The court found that the land acquisitions being carried out did not conform to legal requirements of the Land Acquisition Act, including matters such as the right to judicial review of decisions on compulsory takings, and that the program also did not conform to requirements of the constitution of Zimbabwe (Human Rights Watch 2002b, pp. 14–15). Among other things, the court found that the constitutional requirement of an established program of land reform prior to land confiscation was not met. The judgment declared that

> [T]he settling of people on farms had been entirely haphazard and unlawful: a network of organizations, operating with complete disregard for the law, had been allowed to take over from the Government. War veterans, villagers and unemployed townspeople had simply moved onto farms, encouraged, supported, transported and financed by [ZANU] party officials, public servants, the Central Intelligence Organization and the army.[13]

For devotees of administrative law, the judgment is notable for its invalidation of the law permitting President Mugabe to exercise legislative authority for temporary periods of time, the Presidential Powers (Temporary Measures) Act. The court struck this law down as an unconstitutional delegation of legislative power.

If we stop at this point, it would seem to be a story about the efficacy of the rule of law. The courts stepped in against a very determined government led by a long-time president to protect the rights of property owners and to enforce legal rules. That is the role of courts operating under the rule of law.

But the story goes on. President Mugabe, displeased at the ruling, replaced the Chief Justice of the Supreme Court (whose resignation was widely reported as being forced by Mugabe).[14] The new Chief Justice then named three more new judges to the Supreme Court. With the newly reconstituted bench, the land acquisition case was then reconsidered by five members of

the court. Apparently the Chief Justice selected which five judges would hear the case. This time, the judges decided that the law met both constitutional and other legal standards. The four judges voting to uphold the law were the four new additions to the court (Human Rights Watch 2002b, p. 15). The lone dissenter subsequently stepped down, also reportedly under pressure from the government and its supporters.

Zimbabwe and the rule of law: first cuts

At this juncture, having reached the end, the story looks like a classic breakdown of the rule of law. It is a story about using legal forms to implement the will of the rulers, specifically President Mugabe.

The law prevented Mugabe from doing what he needed to do politically in order to stay in power or at least to be more secure in power. Legal rules prevented him from taking property that belonged to some citizens (the white minority) in order to give it to others (veterans, militia members, government officials, other supporters) who were more critical to Mugabe's political future.

So Mugabe changed the law and he changed the personnel who interpret the law. Mugabe changed the law by using special powers conferred on him to amend the law and by using his parliamentary majority to amend the law, including amendments that purported to apply retroactively (ibid., p. 15). And Mugabe changed the law-readers by using his power to replace judges who opposed him with judges who were more compliant, more willing to accede to his desires. The judges who stepped down were pressured by government officials and by others. The judges were subject not merely to criticism or threats of reduced resources or of having government ignore their rulings. The judges faced death threats, threats they apparently viewed as more than mere rhetoric.

There is little that is more at odds with the rule of law than the use of raw power to bend legal judgments to the will of an individual. There is little more at odds with the rule of law than, in the midst of a legal controversy, changing who reads, who interprets the law in order to change the reading – much less to do so after the law has been read. There is little efficacy to property rights when they can be abrogated so easily to serve the purposes of a ruler. Zimbabwe's treatment of property rights and its treatment of the legal system that defines those rights fails the rule of law because it is a story of the exercise of arbitrary, dictatorial authority.

Arbitrary rule or rule of law?

As obvious as the disparity between the events in Zimbabwe and conformity to the rule of law is – and it should be quite obvious – the line between blatant derogation from the rule of law and conformity to the rule of law,

between behavior consistent with legal protection of property rights and total disregard of those rights, is not so easily drawn on conceptual grounds. The line rests on differences in degree in several respects and, most of all, on the combination of circumstances that free the decisions in Zimbabwe from control by valid, external authority, independent of the identity of the decision maker.

Rule of law and changing law

Consider, for instance, the change in governing law. Property rights will not be secure if the law governing them is subject to change. Much of the rule of law focuses on predictability, and so, too, much of the security for property turns on predictability (see, for instance, Holmes 1897; Hart 1961, pp. 123–8; Fuller 1969, pp. 38–81; Siegan 1997; Barnett 2001, pp. 89–90). Recall Hayek's assertion that the essence of the rule of law is 'that government in all its actions is bound by rules fixed and announced beforehand', and that the rules must 'make it possible to see with fair certainty how the authority will use its coercive powers in given circumstances and to plan one's individual affairs on the basis of this knowledge' (Hayek 1944, p. 80). Even for those who reject the Hayekian approach, predictability has been the touchstone of well-ordered legal systems for a broad array of scholars and judges (Holmes 1897; Fuller 1969). So, changes in law of the sort made in Zimbabwe – to change the procedures to make it easier to take property, to remove the protections against arbitrary selection of the properties to be taken, to eliminate the guarantee of compensation for property acquired by the government – seem incompatible with well-functioning property rights systems.

Legal change under the rule of law

Yet, it is not the mere fact of changes in the law that makes this so. There is no way to bar change in the law or to make property rights absolutely secure against such change. And no legal system has done that.

The United States, which takes pride in its legal system's protection of individual rights, including rights to property, has never proscribed all changes in law that reduce the value of property rights. It has never adopted a blanket prohibition of changes that take away or reduce the value of property rights without compensation. Certainly, evolution of the common law includes decisions that look a great deal like changes in the law that introduce new liabilities, many of which impose burdens on rights to use or dispose of property.[15] Bolder changes in legal rules also take place through legislation or through administrative action. Taxes are raised on some types of property, lowered on others, newly imposed on others; regulation takes away some economic opportunities that existed and imposes new costs on other uses of property. Proposals to implement new taxes on property or to alter the mix of

taxes – lowering income taxes, increasing sales taxes, abolishing or raising or changing rules for estate taxes, eliminating the double taxation of income distributed as corporate dividends – like changes in regulatory regimes can, and often do, significantly alter the value of property rights. US law, however, allows many of these changes to take place without requiring transfer of public funds to those citizens whose property values are diminished.[16]

Some commentators have opined that US law should do more to protect the value of property rights against changes in the law (see, for instance, Epstein 1985; Siegan 1997). But given the number of moving parts in the system of law and governance, almost everyone who looks at the issue recognizes that prohibiting changes in the law that reduce the value of existing rights would be paralysing. It is the functional equivalent of a unanimity rule for governance. However desirable such a rule would be in theory, the costs of implementing it in practice are prohibitive when we move beyond the smallest of groups (see, for instance, Buchanan and Tullock 1962; Mueller 1989, pp. 43–55, 110–11). Among other reasons, such a rule dramatically increases the costs of potential hold-outs, granting individuals strategic power to extract more from fellow citizens than the value they would sincerely place on assent to a given change (Buchanan and Tullock 1962; Epstein 1985). In addition, each hold-out potentially can extract the full social value of a change (Epstein 1985). Although it is not likely that decisions subject to unanimity will face this doomsday scenario, it is quite likely that many of them will be very costly to make under a unanimity rule.

Once the prospect of change without fully unanimous consent is admitted, it is obvious that there will be change that adversely affects individuals and that the harmed individuals will have to bear costs from the change. Unless there is a perfectly error-free way to determine the value of change, positive or negative, and to arrange payments that perfectly adjust for the changes in value experienced by each citizen – a prospect beyond the dream of the most ambitious central planner – there is no way to avoid this outcome.[17]

The prospect of legal change that is permitted and is not limited by a requirement of fully offsetting compensation may be positively beneficial, rather than simply an accommodation to the costliness of centralizing the compensation decision. Individuals acting on their own behalf often are best positioned to make adjustments to the possibility of legal change, including changes that affect property. There is an enormous array of settings in which the value of property rights may be affected and an equally broad variety of ways in which those who hold the rights may assess the risk and calculate the implication for their rights' value. The approach in market economies generally has been to recognize the difficulty of centralizing choices when the welfare effects of a choice are dependent on myriad individual valuation decisions. The dominant strategy in such situations is to set only basic ground

rules and leave each individual free to choose the means of enhancing and protecting his or her investments that seems best suited to that person's individual preferences and expectations.

A regime with little specific legal proscription against – and little specific compensation for effects of – legal change can be assimilated to the approach taken with respect to the general run of social choices in market economies. A now-standard argument regards legal change to be indistinguishable from other changes that affect the value of property rights, such as changes in market conditions, in supply costs, or in consumer tastes (Kaplow 1986). First-party insurance (formal insurance or behavior tantamount to insurance) arguably is better suited to protect against the risk of change than are efforts to constrain the ambit of government action. This is so precisely because it allows each individual to decide how much the insurance is worth and how best to adjust his/her own conduct to limit the risks associated with change. Constraints on legal change are more likely to be blunt instruments, less subject to tailoring to particular circumstances and values.

Restraints on legal change

Legal rules and institutions in the United States, do, however, put limits around how far government can go in changing laws. First-party insurance is only *arguably* better, not certainly better. There are settings in which the change in the law is manifestly subject to the sorts of welfare-reducing manipulations that are inconsistent with rule-of-law values. US law – like the law of other economically advanced democracies – puts both procedural and substantive constraints on such change rather than shifting to the affected parties the burden of insuring against it.

So, for example, changes that affect particular, identifiable individuals, rather than a broad, diffuse class of people, are generally suspect. Often the individuated decisions are retrospective and can impose burdens that are particularly difficult to adjust (Cass 1995; Fuller 1978). But even when prospective, they are associated with a variety of potential decisional biases. For that reason, administrative bodies must give individualized hearings to the affected parties in most settings in which the parties will bear individually differentiated burdens.[18] This requirement has been found in the US Constitution's prohibition of the deprivation of property without due process, as well as in a variety of statutory enactments.[19] The US Constitution also prohibits the national legislature from singling out individuals for special burdens, and most state constitutions in the United States have more general proscriptions against legislation that is aimed at particular individuals rather than at generic classes.[20]

Two other headings relevant here that prevent changes in US law are the 'Contracts Clause' and the 'Takings Clause' of the US Constitution. The first

of these forbids states from impairing the obligation of contract.[21] This clause has not been interpreted as banning without exception any state law that affects existing contractual rights, having historically been seen as permitting the adoption of regulations aimed at protecting the population's health and safety even if the regulations change the existing balance of benefits and burdens under contracts.[22] The concern of the clause's drafters appears to have been legislation aimed specifically at altering contractual commitments, at relieving, for example, debtors of the obligation to pay their debts.[23] Efforts to proscribe that sort of impermissible government law-changing, to circumscribe that set of laws, have been criticized as yielding too modest a constraint (Epstein 1984). Still, the clause does stand as an impediment to a class of possible changes to the law.[24]

The second of these constitutional prohibitions, the Takings Clause, also has a limited application but clearly constrains some government conduct and has enjoyed something of a renaissance in recent years. This clause prohibits government from taking private property for public use without just compensation.[25] Government regulations that address perceived matters of public welfare – from nuisance abatement to protection of historic landmarks – may be upheld despite substantial impact on particular property.[26] This leaves some burdens – sometimes, substantial burdens – to be borne by those whose property is affected. But if government wants to acquire property or if it imposes burdens on it that essentially amount to an acquisition, the Takings Clause requires the government to pay for the privilege.[27] Regulation of property, if it is not to be deemed a confiscation for which compensation is due, must demonstrate a substantial connection between the regulation and its ostensible goal.[28] Even if a regulation meets this test, the regulating jurisdiction cannot avoid paying compensation if the regulation essentially deprives the owner of economically viable use of the property.[29] And the exaction required of a property owner cannot place a disproportionate burden on the property but instead must be reasonable 'both in nature and extent to the impact of the proposed development'.[30]

Putting aside the question whether the requirements are set out in tests that have sufficiently low administrative costs and other error costs to be deemed welfare enhancing overall, the requirements do set out rules that should increase the alignment of government action with social welfare (see Siegan 1997).[31] The connection between means and ends should reduce the prospects for the sorts of exactions that are so obviously welfare reducing that officials would never justify them on their actual grounds. The requirement that exactions must be reasonable in nature and degree relative to the proposed improvement also should increase the probability that the government acts to improve welfare, as that presumably is implicit in the reasonableness judgment. The requirement to pay for regulatory change that effectively

removes all economic value also should promote welfare, as the regulating authority should be willing to make that trade-off if the benefit of the regulation exceeds its cost.[32] The welfare-improving characteristics of the takings rules help constrain changes in ways that also increase predictability and effectively limit official discretion.[33]

These rules do not ensure that no change in the law occurs in the United States, or that all applications of the law are fully anticipated, or that changes in the law or its application do not impose uncompensated harm. The rules do, however, militate in favor of increased restraint on official power and increased predictability in its exercise.

Even more than the legal rules, the barrier to unpredictable action in the United States is the systemic structure that checks government power, that divides authority, that creates governance institutions that make strong, rapid change in law difficult. Lawmaking in the United States is a slow and cumbersome affair. Government power is divided among national, state and local competencies and among a variety of officials. When law cannot be changed quickly, when a change requires consent of large numbers of people, it is more likely that those who will be most affected will be able to anticipate and adjust to the change.

Important constraints on using faster, easier procedures are contained in the US Constitution, and the processes for constitutional change place substantial impediments to easing those constraints. Despite many efforts to amend the Constitution, following the framing era only 17 amendments have been adopted in a span of more than 200 years.[34] The paucity of amendments reflects widespread acceptance of the governance system. That acceptance acts as a check on many potential threats to the system. The commitment to the US governance system is so thoroughly ingrained in the public psyche that even presidents whose political survival is threatened have been unwilling to challenge the system directly (Cass 2001, pp. 34–5).

Although changes in law are accepted in all nations, the commitment to limited and divided power that restrains legal change, and to legal rules that limit the scope and impact of such change on property rights, appears far greater in nations like the United States than in Zimbabwe. The combination of procedural and substantive rules in law-governed nations such as the United States limits the likelihood of legal changes that impose harm on individuals far out of line with what might reasonably have been anticipated. That difference makes change of law in the United States more likely to comport with the rule of law and less likely radically to alter established property rights.

Legal change and land reform

The problem of changing law in ways that affect property rights is especially acute when there is pressure for radical change, such as land reform. Societies, like Zimbabwe, that begin with extreme inequality in distribution of resources, typically land, rarely can address the distributional asymmetries without altering expectations of property rights owners.

Although faulted by the Zimbabwe Supreme Court and others for failing to engage land reform seriously, the genesis for the actions of the Mugabe government lies in the perceived need to spread land ownership more broadly, a special problem for a society that has little wealth apart from land (Mitchell 2001). Mugabe tried to justify the lack of compensation for the land grab by declaring that the black citizens who would now occupy the land collectively represent those from whom whites had stolen the property long ago. His message was not that former owners would get back parcels they had owned – prior to the European colonization of Zimbabwe, there apparently was no developed system of individual land ownership and certainly no recording system that would permit assignment of current land to former owners. Mugabe's assertion, rather, was that the colonization and land division subsequent to colonization constituted theft from the indigenous Zimbabwean population.

When do such general historical grievances provide a basis for undermining established property rights? In the United States there have been at least three occasions for confronting this question. First, there was a protracted, agonizing debate over the treatment of slaves, which were treated as property in the southern states. President Abraham Lincoln had proposed a $400 million fund to compensate former slave owners for the lost value of their slaves, a proposal unanimously opposed by Lincoln's cabinet.[35] Instead, the Thirteenth Amendment to the US Constitution recognized the freedom of former slaves and forbade slavery, without any provision for compensation to those who, in keeping with the laws of their time and place, had regarded human beings as property. A second episode, also unconscionably protracted, addressed claims of indigenous Americans for land taken from them. Although the US government for years was scandalously unfaithful to its treaties with Indian tribes and continues to follow the legal rule that the national government can abrogate aboriginal claims to land without compensation, the government also has continued to provide funds for settling claims of aboriginal title to lands in the eastern United States and to provide property in the West.[36] Moreover, Indian land claims based in deprivation of property for which title was recognized in a valid treaty have been entertained long after the deprivation.[37] Finally, there was the much-needed land reform in Hawaii, where, as the residue of a feudal landholding system, more than 95 per cent of the land was owned by the government or by merely 72 private landown-

ers.[38] In the 1960s, Hawaii passed a land reform law that allowed large landholdings to be broken up, with compensation then paid to the owners for the value of the redistributed land.

The extraordinary injustice of the property holdings in each case – the manifest injustice of slavery, the dispossession and large-scale devastation of aboriginal peoples, the historic anomaly of excessively concentrated landholding in Hawaii – make the claims for redistribution, for alteration of existing rights, compelling; and each seems *sui generis*, separate from other aspects of US law. Had the emancipation of slaves not been concomitant to armed rebellion, there doubtless would have been substantial compensation. Apart from its special nature, the emancipation of slaves is unlikely to create problems with future investment in property and future expectations of the American legal system. It is obviously a very special case. So, too, are our relations with Native Americans. The initial eviction of aboriginal peoples differed from other European colonizations in its attention to the forms of legality, though the claims for compensation from the government later received a more sympathetic hearing than has been common elsewhere. As Alexis de Tocqueville ([1848] 1899, pp. 360–61) observed in the mid-nineteenth century, 'The conduct of the Americans of the United States towards the [Indians] is characterized … by a singular attachment to the formalities of law. … It is impossible to destroy men with more respect for the laws of humanity'.[39] However unjust, there is no current implication for the US rule of law – no one is apt to see the treatment of native peoples today as the bellwether for property protection in general. Finally, the Hawaii experience is a relatively benign and law-bound example of land reform to address historic problems.

Claims for redress of historic grievances are unusual and stretch the bounds of legal systems. Apart from such claims for radical reform, Western legal systems generally create barriers to direct redistribution of property other than through voluntary exchange, even when misbehavior of some sort contributed to an earlier transfer of the property. So, for example, the commercial law protects purchasers of property from various claims of third parties if they have purchased the property in good faith for value (White and Summers 1995, pp. 186–90). This treatment reduces the risk to market transactions, placing the onus for protecting property on the original owner, who is left with a claim for subsequent compensation from the misappropriating party. In cases of serious historical grievance, however, it is understood that different rules may be needed and that, where land redistribution is the issue, full compensation to those whose rights are affected may not be possible within the budget constraints societies face. In such instances, as in Hawaii's example, the requisite effort seems to be to provide an orderly transition and sufficient compensation to reasonably accommodate expectations.

If the Zimbabwe land reform saga seems to augur a more general departure from the rule of law, the defect in Zimbabwe is not in the endeavor to reform land ownership nor is it in the failure to provide compensation fully satisfactory to current owners. The owners might not have foreseen the exact shape of the land reform measures and certainly would have expected some compensation, but they surely anticipated some reform measures. The derogation from rule-of-law values must inhere in some other aspect of the Zimbabwe story.

Legal change and discretionary authority

The process by which the Zimbabwe land reform law was changed appears strongly at odds with legitimacy requisites of the rule of law. The process is not flawed simply because it failed to follow the law of Zimbabwe fully. In many settings, that failure could be cast as one that deprived provisions of the new rules of legality but not of more fundamental legitimacy, though the term 'legitimacy' is sometimes used to connote this more technical aspect of legality. Rather, the process in Zimbabwe is flawed because even the steps taken in conformity to law – such as Mugabe's use of legally conferred emergency powers – confer too much power in too few hands. It is an objection not to the formal *legal validity* of the steps taken by Mugabe but instead to the absence of sufficient governing external authority to make the power exercised by Mugabe law-governed (Cass 2001, pp. 12–18).

This objection is critical to rule of law values, but it is not an easy one to pin down. It is an objection to the scope of discretion given to one person. Too much discretion in any person creates the rule of men, the antithesis of the rule of law. Yet, there is no clear test for the permissible scope of discretion acceptable under the rule of law, and the rule of law necessarily accommodates some scope for official discretion. Differences in judgment about the legitimate scope of discretion that can be vested in official decision makers fuel a good bit of the debate over the meaning of the rule of law and over the conformity of particular legal regimes to it (see, for instance, Dicey 1915; Hart 1961, pp. 119–20, 195–207; Fuller 1969, pp. 86–91; Raz 1979, pp. 206–9).

In even the most rule-of-law-oriented regime, there will be instances of state-sanctioned discretion that seem to be strongly in tension with protection of property rights. The United States, for example, countenances the use of official discretion to order the destruction of property without compensation when it is deemed to be necessary to public safety in times of emergency. A public official can destroy property to create a firebreak or can order animals killed to prevent the spread of disease without the need to offer compensation to the property owner whose investment is sacrificed to the public good (see, for instance, Keeton et al. 1984, p. 146; Christie 1999). Administrative agen-

cies can render decisions on licenses that dramatically reduce the value of the license or that take licenses away from holders who have invested substantially in them without compensating the injured party (Cass 1982, p. 2). Property owners and developers have complained that officials have required them to provide cash payments for new schools, for roads, for water and sewer improvements, for affordable housing projects, to support a council for older citizens, to fund the local park department, or to defray costs of general services.[40] In each case, the assertion by the owner is that some perfectly reasonable use of their property is held hostage to exactions that unreasonably burden the property.

Real property is especially vulnerable to such discretionary exactions. Whether reasonable or not, the quid pro quo for use of the property takes advantage of the property's immobility. Because the owner cannot remove the property to another jurisdiction, local regulators can – subject only to the limits of political possibility – impose conditions up to the value of the improvement. This is, in effect, the mirror image of the holdout problem that justifies forced takings by the state (Epstein 1985). Various forces, such as the effects of repeat play between developers and regulators and competition among jurisdictions for the investments that secure property rights induce, may constrain the exercise of official discretion.[41] But in almost any regime with captive property (property subject to a single regulating authority) and costly monitoring of official behavior, there will be at least some residual discretion. It seems, then, that the existence of discretion, even discretion that can visit substantial harm on property rights owners, is not clear evidence of a regime departing from the rule of law.

Indeed, some discretion may be necessary to support the rule of law. Economic analysis of legal rules suggests that there is an optimal degree of precision that is consistent with social welfare, a degree less than total precision (Ehrlich and Posner 1974; Diver 1983). The point is not simply that precision is costly to achieve. Too much precision also can generate such complex rules that there will be costs to rule application that exceed any gains from the increased precision. Some degree of discretion in rule application, thus, will be welfare enhancing.

The same conclusion must be true in respect of the rule of law. One requisite of the rule of law is predictability in the application of rules based on the rules themselves. Efforts to eliminate all discretion doubtless would produce a body of rules so complex and cumbersome as to make the application of rules *less* predictable, not more so. In fact, overly complex rules may generate more effective discretion than simpler rules overtly conferring a degree of discretion on rule enforcers. The goal for rule-of-law purposes, hence, is not to eliminate discretion but to find a level of discretion that is consistent with predictability and constraint.

In evaluating the experience in Zimbabwe, however, the finer points of this debate need not detain us. The discretion conferred on Mugabe by the Presidential Powers (Temporary Measures) law was virtually unlimited both in the subjects on which Mugabe could legislate and in the nature of the legislation he could enact.[42] And Mugabe used that discretion not to effect a land reform program that would apply neutrally to those who are similarly situated, instead tilting for and against favored and disfavored groups. Wherever the line lies between permitted discretion cabined by law, on the one hand, and discretion so broad as to erode a sense of legitimate governance under law, on the other, that line places the authority conferred on Mugabe outside the bounds of the rule of law.

Changing personnel

As with the other issues that draw Zimbabwe's actions into question, the change in the identity of the judges who interpret the law must be seen as something that violates the rule of law because of its particulars, not because any change – even any change aimed at altering the law – is illegitimate. The rule-of-law requirement of principled predictability means that the identity of the individuals who interpret the law should not matter. The legal rules themselves should have sufficient specificity and sufficient binding force to allow prediction of their application without knowledge of the particular individual who will interpret the rules (Dorf 1995; Cass 2001, pp. 16–18). If that is not so, not only will the system lack principled predictability, it also will lack the legitimacy that comes with adherence to external authority (Schauer 1991; Cass 2001, pp. 18–19). If those who interpret and apply the law have sufficient discretionary authority that their identity is critical to the outcome, it can hardly be said that the laws rule, even if it can be said that the nominal rulers – the political officials holding primary positions of power and authority in government – do not.

This much is fairly common ground, a basis for agreement among those who favor relatively expansive modes of judicial interpretation and those who favor relatively constrained modes, for those who believe that the laws can and do provide determinate structure for judicial decisions and for skeptics on those matter (Singer 1984; Dworkin 1986; Scalia 1989; Easterbrook 1991; Dorf 1995). What is debated quite vigorously is how much the laws actually do constrain and how much freedom judges actually enjoy over the disposition of issues that come before them. It is a debate about how much judges matter, and to some degree *which* judges matter.

In the United States, this debate has been quite visible in the public domain recently. It may be especially important not only because so many matters come before courts, but also because courts exercise somewhat broader powers in America than in most other nations (Posner 1996). A subcommittee of the

US Senate has held several hearings on the extent to which ideology does and should matter in the work of judges and in the selection of judges. The focus of those hearings, as of most debates on these issues, was on judges on the nation's highest appellate courts. These courts decide a tiny fraction of the judicial cases filed in America and address legal issues that are least governed by fully articulated external authority. Plainly, for these courts more than the general run of 'ordinary' courts, the identity of judges matters, as the greater openness of governing law increases the influence of all manner of other considerations that doubtless vary across individuals (Cass 2001, pp. 62–71, 84–97). The officials in the executive branch will care who these judges are, and so, too, will members of the opposition party. Each will endeavor to appoint individuals who will tilt toward preferred outcomes on important, contested matters.

On the surface this endeavor looks to be a direct contradiction to the rule of law requirement that the legal authority, rather than the law interpreter, should govern. So, what makes the endeavor to select judges sympathetic to particular views consistent with the rule of law? What, in other words, makes the United States in this respect unlike Zimbabwe?

First, as with the concentration of discretionary power over legislation under the Presidential Powers law, the selection of judges in Zimbabwe looks to be much more subject to the will of a single individual than is the case in the United States, where the President appoints members of the federal government's judiciary, but he can do so *only* if he is able to secure agreement from the US Senate, agreement that has been famously lacking in numerous instances. The credible threat of rejection colors the nomination and appointment process. In Zimbabwe, the President selects judges and is only required to consult with a judicial service commission, not to secure its approval.[43] If the commission disapproves, the President informs parliament, but there is no provision for parliamentary disapproval of the selection.[44]

Beyond the confirmation of individual judges, in the United States the Congress prescribes the number of judges authorized for the various federal courts. Although presidents have from time to time proposed changes to the number of judges – most notably President Franklin Roosevelt's plan to add members to the Supreme Court in order to facilitate its acceptance of key aspects of his program (Leuchtenberg 1966) – politically inspired changes are rarely adopted. Roosevelt's plan, for example, was not enacted.

In Zimbabwe, the number of judges, even of the Supreme Court, is subject to fairly easy manipulation. The President can designate the number of judges he 'deems necessary' to sit on the Supreme Court.[45] And if the Chief Justice decides that more are needed (not a defined term in the constitution) on a temporary basis (also not defined), he can appoint additional judges to the Supreme Court for a term of his choosing.[46] In combination, the President

and a chief justice sympathetic to him can have a powerful impact on the size and shape of Zimbabwe's highest court.

Without effective opposition, the President of Zimbabwe can select judges who are most likely to be cooperative on the matters of greatest immediate interest to him, and with the added ability to set the judges' terms of office can also replace them later to secure desired outcomes on other matters. If he needs additional judges to secure a majority that bypasses the entire sitting Supreme Court bench, he can add the requisite number. That is what occurred in the land reform saga: the selection of enough cooperative judges to secure a particular result in a particular case of importance to the nation's chief executive. This is not the overlay of politics on a judicial decision process that is otherwise insulated from politics. It is not the subjection of judicial selection to the pull and tug of competing political forces. Instead, it is the substitution of concentrated political power for judicial independence.

Second, perhaps in response to the pressure that can be brought to bear by the President, the process of judging in Zimbabwe does not at the end of the day look so well insulated from political influence as in the United States. Notwithstanding the very public criticisms of the US appellate courts for decisions characterized as overly political, US courts, including appellate courts, behave in a manner highly predictable from the governing law. Courts show precious little influence of politics – and US courts, including those thought to be most subject to ideological influence, are amazingly often in accord regardless of the political background of the deciding judges (see Cass 2001).[47] Efforts to influence selection of judges to courts that operate strongly under the rule of law necessarily look less threatening than efforts to select judges whose behavior post-selection is far more apt to reflect the views of the selecting authority.

There are, of course, contrary assertions about the roles of politics and political pressure on the US judiciary. Some scholars view the US judiciary as quite affected by political considerations (Revesz 1997). Episodes such as the change in law during the New Deal era of the 1930s and 1940s, when the US Supreme Court first struck down economic regulation and then in short order upheld similar legislation, are explained as responses to political pressure, specifically in this case to the Roosevelt court-packing plan. The change of a single vote, shifting a 5–4 majority against legislation's constitutionality to a 5–4 majority in favor, could be a simple response to political pressure.[48] But it also could be the product of a single justice thinking more about a difficult, closely contested issue, to the influence of additional conversations, to the impact of seeing additional fact patterns – to any of the myriad non-political factors that can affect decisions where legal texts are relatively open. Doubtless, considerations external to the narrowest view of legal texts influence some decisions even in the most law-bound systems. But a fair view of

the legal system in the United States, as in other Western democracies, still must conclude that there is quite strong insulation of judging from politics and precious little evidence of any direct influence from political officials. This stands in marked contrast to the situation in Zimbabwe.

In sum, the reason why the steps taken in Zimbabwe to undermine property rights and to alter legal rules violate rule-of-law values is not that there is a generic conflict with the type of action – with efforts to change legal rules, to redistribute property, to provide discretion to official decision makers, or to appoint judges congenial to the executive. The steps taken violate rule-of-law values because each oversteps the bounds within which such conduct may be said to permit reasonable certainty about the law and to permit legal rules to operate free of undue control by current rulers.

Conclusion

Governance systems that limit official discretion to impair property rights, that have institutions and rules that provide clear definition to property rights and that provide predictable and consistent applications of those rights, will accord with the rule of law and generally will also advance social welfare. Some systems will depart quite evidently from this pattern, to the detriment of those societies, allowing too ready changes in law at the discretion of too few officials, too unconstrained by law. But differences between the good and the bad will not be drawn along simple, discrete lines. The systems most consistent with the rule of law will not be able effectively to bar all changes in the law or to eliminate official discretion. Instead, those systems will limit the avenues for change and the ambit of discretion in ways that make property more secure and impositions on it more predictable without reference to the identity of the individual official enforcing the law or the individual property owners subject to it. Those will be the legal analogues of Tolstoy's happy families.

Notes

* Thanks for helpful comments (and absolution for remaining errors) are due to Randy Barnett, Jack Beermann, Bob Bone, Ward Farnsworth, Tamar Frankel, Pnina Lahav, Gary Lawson, David Lyons, Allan Macurdy, Bob Seidman, Ken Simons, and to participants in the Boston University Law Faculty Workshop.

1. See for instance *Goldberg v. Kelly*, 387 U.S. 254 (1970); *Board of Regents v. Roth*, 408 U.S. 564 (1972).
2. This line differs from that drawn by scholars who distinguish property rights from other rights along remedial lines, with property rights enforceable through injunctions and other rights providing only an entitlement to money damages. See, among others, Calabresi and Melamed (1972); Bebchuk (2001).
3. In federal systems with multiple levels of government, more than one government may affect property. But the significant point is that the property is not movable to avoid the edict of the relevant authority.
4. This assumes that regulation imposes greater burdens on those whose property is affected

than the benefits conferred on them through regulation. That is not the same as assuming that regulation has negative net welfare effects for society as a whole, though many commentators have made that case for economically advanced democracies as well.

5. See, among others, Crain and Hopkins (2001), who suggests a figure in excess of $840 billion for the United States, and Hopkins (1994).
6. Tapan Munroe, 'Economies of U.S. and Japan', *Contra Costa Times*, 22 September 2002.
7. Securities and Exchange Commission (2002), 'Table 15: Value of Stocks Listed on Exchanges', *SEC Annual Report 2001*, Washington, DC: US Government Printing Office, p. 169 (putting the figure at more than $11.7 trillion). This figure was from 2000, and there has been substantial erosion in the US stock markets since then, making the figure unreliable as a precise measure of current worth. Still, the point in the text holds: the value of US stockholdings is very substantial. Further, the stock figure does not include the value of the very large number of businesses that do not have publicly traded shares.
8. Constitution of Massachusetts, Declaration of Rights, article 30 (1780, authored by John Adams); see also Hume ([1742] 1985, p. 94). A similar phrase is found in Harrington ([1656] 1992).
9. Constitution of Zimbabwe, § 16A. Human Rights Watch (2002a, p. 10); 'Parliament Approves Land Reform Bill', *World Markets Analysis*, 7 April 2000.
10. See 'Zimbabweans Reclaim Birthright', *The Herald* (Harare), 10 December 2002. Although it is the only figure I have found for this time frame, it is not entirely clear that it is a credible one. Other sources list under 4.5 million hectares as resettled a year earlier with a little more than 9 million hectares scheduled by the government for acquisition. See Human Rights Watch (2002a, p. 11). It would seem unlikely that 80 per cent of the land potentially to be acquired would so quickly have passed out of the control of the land owners. It is possible, however, that the figure is high because it includes land that is occupied by squatters even though not formally listed for potential acquisition by government. See Human Rights Watch (2002a, p. 13).
11. 'The Great Terrain Robbery', *The Economist*, 17 August 2001, p. 37; Human Rights Watch (2002a).
12. See, for instance, *The Economist* (see note 11); PRS Group (1 December 2001), *Zimbabwe: Politics*, East Syracuse, NY: PRS Group; Daniel Foggo, 'VC Hero Forced Off Land by Mugabe Thugs', *Sunday Telegraph*, 8 December 2002, p. 3. See also *Commercial Farmers Union v. Minister of Lands, Agriculture & Resettlement*, Zimbabwe Supreme Court, Judgment No. SC-132/2000, 21 December 2000.
13. *Commercial Farmers Union v. Minister of Lands, Agriculture & Resettlement* (see note 12).
14. See, for example, PRS Group (1 September 2002), *Zimbabwe: Country Highlights*, East Syracuse, NY: PRS Group; 'Magistrates Put Under Pressure in Land Rulings', *Zimbabwe Independent*, 20 September 2002. The former Chief Justice, Anthony Gubbay, apparently had received numerous death threats following the ruling on land reform. See, for instance, Human Rights Watch (2002b, p. 15).
15. See, for example, *Henningsen v. Bloomfield Motors*, 32 NJ 258, 161 A.2d 69 (1960).
16. See, for example, *Penn Central Transportation Co. v. New York City*, 438, U.S. 104 (1978).
17. This has been a staple part of the long-running argument over use of Kaldor–Hicks-type efficiency criteria for governmental decisions.
18. See, for example, *Goldberg v. Kelly*, 387 U.S. 254 (1970).
19. Compare *Londoner v. Denver*, 210 U.S. 373 (1908), with *Bi-Metallic Investment Co. v. State Board of Equalization*, 239 U.S. 441 (1915).
20. US Constitution, Article I, § 9, clause 3; Constitution of Colorado, Article V, § 25; Constitution of Georgia, Article 3, § 6, ¶ 4; see *United States v. Lovett*, 328 U.S. 303 (1946).
21. US Constitution, Article I, §10, clause 1.
22. See, for example, Manigault Springs, 199 U.S. 473 (1905). For a similar construction of the due process clause of the US Constitution, see *Mugler v. Kansas*, 123 U.S. 623 (1887).

23. See, for example, *Fletcher v. Peck*, 10 U.S. 87 (1810); *Sturges v. Crowninshield*, 17 U.S. 122 (1819).
24. See, for example, *Ruckelshaus v. Monsanto Co.*, 467 U.S. 986 (1984).
25. US Constitution, Amendment V, clause 5.
26. See, for example, *Penn Central Transportation Co. v. New York City*, 438, U.S. 104 (1978).
27. See, for example, *Nollan v. California Coastal Commission*, 483 U.S. 825 (1987); *Loretto v. Teleprompter*, 458 U.S. 419 (1982).
28. *Nollan v. California Coastal Commission*, 483 U.S. 825 (1987).
29. *Lucas v. South Carolina Coastal Council*, 505 U.S.1003 (1992).
30. *Dolan v. City of Tigard*, 512 U.S. 344 (1994).
31. Komesar (2002), argues that the tests correct for some types of welfare-reducing incentive problems but exacerbate others.
32. Of course, that statement puts a bit of a rabbit in the hat by assuming that regulators are motivated to improve overall social welfare – an assumption that is routinely called into question. The literature on public choice (among other writings) suggests numerous ways in which government officials' incentives diverge from pursuit of overall social welfare. See, for instance, Buchanan and Tullock (1962); Olson (1965); Shepsle and Bonchek (1997).
33. In part, the rules increase congruence with the rule of law by making it more likely that official decisions will be reasonable. Reasonableness is a relevant concern when there is not a clearly formulated, formally promulgated regulation. See Fuller (1969, p. 138); Schauer (1991); Cass (2001, pp. 12–15). It also is relevant to interpretation of formally promulgated regulations, where a construction that produces unreasonable results could lack sufficient predictability to meet rule of law requisites. See Fuller (1969) and Schauer (1991, pp. 222–3).
34. These amendments primarily have expanded the democratic franchise, filled in details for government continuity, or granted rights against government power to groups that had been especially vulnerable.
35. See Jay Winik (2001), *April 1865: The Month That Saved America*, New York: Perennial, p. 34. Lincoln previously had by executive order decreed the freedom of slaves in states that were engaged in armed rebellion against the federal government.
36. See, for example, Florida Indian Claims Settlement Act, 25 United States Code § 1741 et seq.; Maine Indian Claims Settlement Act, 25 United States Code § 1721 et seq.; Rhode Island Indian Claims Settlement Act, 25 United States Code § 1701 et seq.; Canby (1998, pp. 343–64).
37. See *United States v. Sioux Nation*, 448 U.S. 371 (1980); Canby (1998, pp. 352–3).
38. See *Hawaii Housing Authority v. Midkiff*, 467 U.S. 229 (1984).
39. I am indebted to Allan Macurdy for bringing this quotation to my attention.
40. See, for instance, Maryls Duran, 'Growing Pains: As Rural Areas Expand, Schools Feel the Crunch', *Denver Rocky Mountain News*, 28 November 1999, p. 48A; Miguel Navrot, 'Developer Questions City Action', *Albuquerque Journal*, 30 September 2000, p. 1; Mia Taylor, 'Builders Often Face Questionable Requests', *Patriot Ledger*, 25 September 1999, p. 1; Mark Vosburgh and Christopher Quinn, 'Developer Says Withholding Building Permits Is Extortion', *Plain Dealer*, 22 January 2002, p. B2; Lesley Wright and Anne Ryman, 'Scottsdale May Force Builder, Schools Talks', *Arizona Republic*, 6 August 2002, p. 1B.
41. Repeat play might moderate or exacerbate regulators' incentives to impose exactions. Competition among jurisdictions should work strictly to moderate those incentives, but just how such competition works is a subject of some dispute. See, for instance, Tiebout (1956); Buchanan and Goetz (1972).
42. Presidential Powers (Temporary Measures) Act, Zimbabwe, Chapter 10:20.
43. Constitution of Zimbabwe, § 84 (1).
44. Ibid., § 84 (2).
45. Ibid., § 80 (2)(b).
46. Ibid., § 80 (3).
47. Ronald A. Cass (September 2002), *The DC Circuit: Considering Balance on the Nation's*

Second-Highest Court, statement to Subcommittee on Administrative Oversight and the Courts of Senate Committee on the Judiciary.

48. Although President Roosevelt complained about the wrongness of a 5–4 decision invalidating legislation, some of the critical decisions prior to the court-packing plan were 9–0 against the legislation while others were 8–1 in favor. Much attention afterward focused on a single 5–4 decision upholding a key piece of New Deal legislation.

References

Barnett, Randy (2001), *The Structure of Liberty: Justice and the Rule of Law*, Oxford: Clarendon.

Barzel, Yoram (1997), *Economic Analysis of Property Rights*, 2nd edn, Cambridge, UK and New York: Cambridge University Press.

Bebchuk, L.A. (2001), 'Property rights and liability rules: the *ex ante* view of the cathedral', *Michigan Law Review*, **100** (3), 601–39.

Berry, Christopher (1980), 'Two replies to Locke – Hume and Hegel', in J. Roland Pennock and John W. Chapman (eds), *NOMOS XXII: Property*, New York: New York University Press, pp. 89–99.

Buchanan, J.M. and Ch. J. Goetz (1972), 'Efficiency limits of fiscal mobility: an assessment of the Tiebout model', *Journal of Public Economics*, **1** (1), 25–43.

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

Calabresi, G. and D. Melamed (1972), 'Property rules, liability rules and inalienability: one view of the cathedral', *Harvard Law Review*, **85** (6), 1089–128.

Canby, Jr., William C. (1998), *American Indian Law*, 3rd edn, St. Paul, MN: West Publishing.

Cass, R.A. (1982), 'RKO: a special kind of lottery', *Media Law Notes*, **9** (4).

Cass, R.A. (1995), 'Judging: norms and incentives of retrospective decision-making', *Boston University Law Review*, **75** (4), 942–96.

Cass, Ronald A. (2001), *The Rule of Law in America*, Baltimore, MD: Johns Hopkins University Press.

Christie, G.C. (1999), 'The defense of necessity considered from the legal and moral points of view', *Duke Law Journal*, **48** (5), 975–1042.

Christman, John (1994), *The Myth of Property*, New York: Oxford University Press.

Crain, Mark W. and Thomas D. Hopkins (2001), *The Impact of Regulatory Costs on Small Firms*, July, Washington, DC: US Small Business Administration.

de Tocqueville, Alexis (1848), *Democracy in America*, reprinted in H. Reeve (trans.) (1899), London, UK and New York: Longmans.

Demsetz, H. (1967), 'Toward a theory of property rights', *American Economic Review Papers and Proceedings*, **57** (2), 347–59.

Dicey, Albert V. (1915), *An Introduction to the Study of the Law of the Constitution*, 8th edn, London: Macmillan.

Diver, C.S. (1983), 'The optimal precision of administrative rules', *Yale Law Journal*, **93** (1), 65–109.

Donohue, Jr., Charles R. (1980), 'The future of the concept of property predicted from its past', in J. Roland Pennock and John W. Chapman (eds), *NOMOS XXII: Property*, New York: New York University Press, pp. 28–58.

Dorf, M. (1995), 'Prediction and the rule of law', *UCLA Law Review*, **42** (3), 651–715.

Dworkin, Ronald (1986), *Law's Empire*, Cambridge, MA: Belknap Press.

Easterbrook, F.H. (1991), 'What's so special about judges?', *Colorado Law Review*, **61** (4), 773–82.

Ehrlich, I. and R.A. Posner (1974), 'An economic analysis of legal rulemaking', *Journal of Legal Studies*, **3** (1), 257–86.

Epstein, R.A. (1984), 'Toward a revitalization of the Contracts Clause', *University of Chicago Law Review*, **51** (3), 703–51.

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

Fuller, Lon (1969), *The Morality of Law*, rev. edn, New Haven, CT: Yale University Press.

Fuller, L. (1978), 'The forms and limits of adjudication', *Harvard Law Review*, **92** (2), 353–409.
Harrington, James (1656), *Commonwealth of Oceana*, reprinted in J.G.A. Pocock (ed.) (1992), Cambridge and New York: Cambridge University Press.
Hart, H.L.A. (1961), *The Concept of Law*, Oxford: Oxford University Press.
Harvey, B.W. (1961), 'The rule of law in historical perspective', *Michigan Law Review*, **59** (4), 487–500.
Hayek, Friedrich A. (1944), *The Road to Serfdom*, Chicago: University of Chicago Press.
Hayek, Friedrich A. (1960), *The Constitution of Liberty*, Chicago: University of Chicago Press.
Holmes, Jr., O.W. (1897), 'The path of the law', *Harvard Law Review*, **10** (8), 457–78.
Hopkins, Th. D. (1994), 'The costs of federal regulation', *Journal of Regulation and Social Costs*, **2** (1), 5–31.
Human Rights Watch (2002a), 'Fast track land reform in Zimbabwe', *Human Rights Watch A*, **14** (1).
Human Rights Watch (2002b), *Submission to Commonwealth Ministerial Action Group*, January, New York: Human Rights Watch.
Hume, David (1742), *Essays, Moral, Political, and Literary*, reprinted Eugene F. Miller (ed.) (1985), Indianapolis, IN: Liberty Classics.
Kaplow, L. (1986), 'An economic analysis of legal transitions', *Harvard Law Review*, **99** (3), 509–617.
Keeton, W. Page, Dan B. Dobbs, Robert E. Keeton and David G. Owen (1984), *Prosser and Keeton on the Law of Torts*, 5th edn, St. Paul, MN: West Publishing.
Komesar, Neil (2002), *Law's Limits*, Cambridge and New York: Cambridge University Press.
Leuchtenberg, William E. (1966), 'The origins of Franklin D. Roosevelt's "Court-Packing" Plan', *Supreme Court Review*, **1966**, 347–400.
Merrill, Th. W. and H.E. Smith (2001), 'What happened to property in law and economics?', *Yale Law Journal*, **111** (2), 357–98.
Mitchell, T.W. (2001), 'The land crisis in Zimbabwe: getting beyond the myopic focus on black and white', *Indiana International and Comparative Law Review*, **11** (3), 587–603.
Mueller, Dennis E. (1989), *Public Choice II*, Cambridge, UK: Cambridge University Press.
Murphy, Liam and Thomas Nagle (2002), *The Myth of Ownership*, Oxford and New York: Oxford University Press.
Nozick, Robert (1974), *Anarchy, State, and Utopia*, New York: Basic Books.
O'Driscoll, Jr., Gerald P., Kim R. Holmes and Melanie Kirkpatrick (2001), *Index of Economic Freedom*, Washington, DC: Heritage Foundation.
Oakeshott, Michael (1983), 'The rule of law', in Michael Oakeshott, *On History and Other Essays*, Totowa, NJ: Barnes & Noble, pp. 157–64.
Olson, Mancur (1965), *The Logic of Collective Action*, Cambridge, MA: Harvard University Press.
Pipes, Richard (1999), *Property and Freedom*, New York: Alfred A. Knopf.
Posner, Richard A. (1996), *Law and Legal Theory in England and America*, Oxford and New York: Clarendon Press.
Rawls, John (1971), *A Theory of Justice*, Cambridge, MA: Harvard University Press.
Raz, Joseph (1979), *The Authority of Law: Essays on Law and Morality*, Oxford: Clarendon Press.
Reich, Ch. (1964), 'The new property', *Yale Law Journal*, **73** (5), 733–87.
Revesz, R.L. (1997), 'Environmental regulation, ideology, and the D.C. circuit', *Virginia Law Review*, **83** (8), 1717–72.
Scalia, A. (1989), 'The rule of law as a law of rules', *University of Chicago Law Review*, **56** (4), 1175–88.
Schauer, Frederick (1991), *Playing by the Rules: A Philosophical Examination of Rule-Based Decision Making in Law and in Life*, Oxford: Clarendon.
Shepsle, Kenneth A. and Mark S. Bonchek (1997), *Analyzing Politics: Rationality, Behavior, and Institutions*, New York: W.W. Norton.
Siegan, Bernard (1997), *Property and Freedom: The Constitution, the Courts, and Land-Use Regulation*, New Brunswick, NJ: Transaction Publishers.

Singer, J. (1984), 'The player and the cards: nihilism and legal theory', *Yale Law Journal*, **94** (1), 1–70.
Singer, J.W. and J.M. Beermann (1993), 'The social origins of property', *Canadian Journal of Law and Jurisprudence*, **6** (2), 217–48.
Tiebout, Ch. M. (1956), 'A pure theory of local expenditures', *Journal of Political Economy*, **64** (5), 416–24.
White, James J. and Robert S. Summers (1995), *Uniform Commercial Code*, vol. 1, 4th edn, St. Paul, MN: West Publishing.

PART III

CURRENT ISSUES FROM A PROPERTY RIGHTS PERSPECTIVE

11 Are property rights relevant for development economics? On the dangers of Western constructivism

*Enrico Colombatto**

On the notion of economics

Modern economic analysis is usually conceived in two different ways. Some believe that economists should study how given policy goals can be attained. In particular, when mutually inconsistent objectives are proposed, social scientists should describe the implicit trade-offs and suggest solutions, possibly by referring to targets of a higher order. The main actors are thus the policy maker, who defines the objectives, and the social engineer, who designs the course of action consistent with the assigned target.

Contrary to this widely shared (mainstream) view, the subjectivist school claims that all social sciences should be dealing with the way individuals behave and interact in order to pursue their own aspirations. The search for a social goal should thus be ignored, for it would necessarily be based on arbitrary utilitarian hypotheses, thereby leading to useless or even harmful results. In particular, (free-market) economists should focus on the analysis of human conduct in the presence of scarcity constraints, whereby individuals have to choose what to sacrifice in order to meet their preferences or – more precisely – to increase their satisfaction.

The differences between these two views have important consequences from the moral, positive and normative standpoints. They also define the different roles of property rights within the context of growth and development. The first part of this chapter is thus devoted to recalling the ethical validation of a society based on private property rights principles and draws some implications for growth and development policies from a free-market perspective. By contrast, the second part focuses on the discretionary nature of the property rights systems associated with the orthodox views, be such views a matter of central planning or of technocratic design. Although the importance of property rights as the engine of growth remains beyond dispute, most of this chapter tries to show that the crucial issue is not so much the definition of the allegedly 'optimal' property rights system, as the understanding of the ideological elements that justify property rights in the first place. Thus, unless one clearly perceives the nature of the ideological structure that legitimises the rules of the game in a given society, it is virtually

impossible to understand why growth-conducive property rights systems fail to emerge and be accepted. In other words, property rights are certainly necessary for growth. However, getting them right is not enough.

Morality, rules and policy making in free-market economics

Morality is probably the key concept. It is usually understood as a set of behavioural rules consistent with the ethical standards shared by the large majority of the individuals forming a community. Therefore, morality changes over time and across communities and defines the patterns of social cooperation.[1] For instance, today torture or capital punishment are morally unacceptable by most people in Western Europe, but under given circumstances they are deemed appropriate in other parts of the world. Similarly, in many countries breach of contract in the workplace is often standard practice if committed by the employee, but it becomes immoral if committed by the employer. Tolerance with respect to theft or fraud is also highly variable across countries and across time.

Indeed, one can probably maintain that such sets of shared subjective values define the communities themselves. From a historical perspective, that was actually the case in much of Europe for a very long time – at least until the sixteenth century. Until then, individuals were first of all Christians (and also Orthodox, Catholics, Reformed), Muslims, Pagans, Jews and so on. Political boundaries were porous and had little to do with national belonging, fatherland and special bonds to countrymen. Loyalty to a ruler was based on the prevailing Christian morality, which prescribed little more than vague obedience to an élite designated by divine power to administer justice and possibly provide protection against external attacks, too.[2]

It is manifest that this approach to morality does not imply any obligation for individuals, who should then be free to pursue their own preferences as long as they do not opt for violence[3] in order to limit other people's freedom to choose. Of course, agents do not necessarily operate in a society without rules. But such rules do not need a forced-upon, predefined moral framework.[4] On the other hand, they require compliance with the principles of human dignity, which cannot be separated from the notion of private property rights, and from people's right to subdue nature. These are at the core of what is usually known as Western culture. Hence, the sacredness of individual freedom, the duty to accept individual responsibility and the inviolability of property rights both over the fruits of one's own labour and efforts and over natural resources not previously appropriated by others. Given these fundamentals, which are rooted in the so-called 'natural law', all other rules that discipline human action should be the result of an explicit contract, whereby individuals voluntarily accept to restrict their behaviour, if they wish to do so.

Contrary to the orthodox approach, where morality is a set of collective goals defined through a process of democratic selection, the moral fundamentals of free-market economics set severe limits to political action, since in this latter case the politician cannot interfere with individual liberty and property rights. He/she can only protect and enforce them; and loses legitimacy when his/her behaviour is at odds with the natural order (see also Hoppe, Chapter 2 this volume).

One can thus conclude that the issues of development and growth[5] in a free-market context in fact are non-issues, for development is hampered whenever the institutional framework restricts economic freedom. Similarly, growth is simply a matter of personal choice, for it reflects the desire to sacrifice leisure and engage in productive activities, given the opportunities available. Surely, there is no need for growth strategies, development policies or grand projects of any kind, but only for an appreciation of the protection and enforcement of property rights in a given institutional framework (which is far from being an easy task, anyway).

In other words, from the free-market standpoint an undeveloped country is one where property rights are not enforced. Per capita GDP plays a role only because this variable is often correlated with the lack of economic freedom. Thus, it provides a clue as to where property rights are more likely to be violated. And inequalities play no role at all, either, unless they originate from arbitrary rules of the game being imposed upon various layers of the population according to religion, race, tribal origin, gender, passport and so on.

Property rights and orthodox development economics

As hinted in the preceding paragraphs, the moral foundations of mainstream economics are not based on subjectivism. Instead, good or bad orthodox policies are defined according to their perceived ability to attain shared goals, rather than to enhance individual preferences. Not surprisingly, economists turn to statistics, or devote their energies to studying political expediency and pragmatism, so as to develop recipes for effective policy making and methods to maximise consensus. Somewhat paradoxically, the repeated efforts to shape economics as a hard science have transformed many scholars into instruments serving a very soft discipline – politics in redistributive democracies. Property rights structures are thus a tool of policy making. As a result, they become subject to political or technocratic discretion when it comes to their definition, their assignment or reassignment, and their enforcement. Mainstream development economics is no exception.

Substituting planning for private property rights

Two distinct periods have characterised traditional post-Second World War development economics.[6] The first coincided with widespread planning am-

bitions and dominated until the early 1970s, possibly a little longer. Despite the disappointing results of pervasive state intervention in the 1930s, both the academic community and public opinion were impressed by the alleged economic successes of the Soviet bloc and eager to reproduce at least some features of the central-planning process. In particular, private property would be accepted, but considered to be subject to aims and principles of a higher order. Even today, the same principle would apply to situations where individuals exchanging property rights are not able or willing to take into account the full implication of their decisions. Of course, this falls short of perfection and can be interpreted as an economic inefficiency. But since this is always the case in any human activity, assessing when such inefficiencies are intolerable has become a matter of discretionary evaluation by the policy maker, the legislator and the judiciary.

The undeveloped world turned out to be an almost ideal target for greenfield experiments in economic policy. In some cases development economics focused on how property rights were to be redefined and transferred to the state, that is, to politicians and various layers of the bureaucracy. In other cases property rights were weakened and made subject to arbitrary intervention by the state, rather than transferred. By and large, however, the assignment, exchange and enforcement of property rights were designed in order to suit political priorities, rather than individual economic needs and preferences.

Not surprisingly, all efforts to reproduce Western or even Soviet economic results failed miserably. In most cases, centralised investment policies and easy credit led to cathedrals in the desert, capital outflows and ultimately heavy debt. Similarly, educational policies led to élite schools for the top politicians, bureaucrats and their clients.[7] Whereas technological transfer actually meant well-remunerated jobs for international professionals and bureaucrats, to be paid for by international agencies.

Surely, property rights were not entirely destroyed. The tragedy of the commons as traditionally understood was not necessarily a distinctive feature of the undeveloped areas. Rather, rights were *de facto* appropriated by the political class and subsequently reassigned, sometimes for limited periods of time, sometimes with no guarantee of enforcement. In many situations this explains the frequent outbreaks of violence, as competing pressure groups tried to appropriate and enforce rights on resources; or to influence their assignment by the political élites and thus acquire some kind of legitimacy *vis-à-vis* third parties, including the international community.

From planning to technocratic utopia

Different editions of the neoclassical paradigm have dominated the second period of orthodox development economics. They all tried to provide explanations for the failures of the previous decades and they all shared the view

that property rights assignments were a matter of technocratic intervention. According to a naive version, economic backwardness was attributed to bad policy making, due to poorly qualified decision makers. Consistent with this view, most recipes aimed at (i) providing expert economic advice and/or a justification to carry out allegedly good economic policy even if resisted by public opinion,[8] (ii) offering financial support in order to enhance political feasibility, that is, to reduce the cost of transforming a rent-seeking society into a free-market economy, (iii) providing financing and know-how so as to realise adequate infrastructure and public goods, without which a market economy cannot work properly, and (iv) introducing democratic political institutions. These measures have not made a great difference to the economies of the undeveloped world. Most of them remained poor. Indeed, many countries that did well in the 1970s and the 1980s followed other avenues and would hardly figure among the suitable case studies for neoclassical textbooks.

After being severely criticised, this naive version has gradually been integrated with another set of prescriptions and transformed into a new story, whereby the influence of the institutional school of thought is apparent. It argues that growth and development are indeed a matter of appropriate institutions (rules of the game), and that the quality of institutions ultimately depends on the definition, exchange and enforcement of property rights. Hence, new development policies have been focusing on the creation of a system of political rules and organisations that replicates – by and large – those currently in place in the developed world. The importance of property rights is now fully acknowledged. Still, its assignment and enforcement remains a top-down process, whereby a local politician or an international technocrat suggests and enforces the most suitable institutional arrangements.[9] As will be discussed at length shortly, this represents a crucial weakness and seriously undermines the traditional property rights view of economic development.

Where do we stand?

There is little doubt that the institutional commitment shown by much of the orthodox economic profession is not going very far. Under these circumstances, private property is not the foundation of human dignity and societal interaction, but rather a tool for policy making. Indeed much of the debate in the recent past has concentrated on defining the extent to which (private) property can be violated, what kind of rules of the higher order are needed in order to limit discretion, how such rules can be enforced. Similarly, a considerable part of the 'political economy' literature has been devoted to explaining which constitutional project stands the best chance of reducing political instability, social tensions (if not civil war), misbehaviour and discretion by

bureaucrats and policy makers (for example, corruption). While hundreds of experts spend their time teaching the virtues of market economics and fair income distribution.

In fact, failures have been caused neither by ignorance about the virtues of property rights economics, nor by the boundless and brutal greed of shortsighted but rational autocrats with an interest in crushing opportunities for development. There is no doubt that the persistent failure to grow owes a great deal to poor rules of the game and unsatisfactory property rights arrangements. But at least equally important is the interaction between deeply rooted informal institutions, old cultural traits, new and possibly disruptive ideological features. Enforcing property rights from outside will not do. For the outlook for growth will remain gloomy as long as the ability to create and seize opportunities for self-enhancement is being hampered. In other words, it is here maintained (i) that property rights cannot be imposed by fiat, (ii) that in order to understand what kind of property rights structure is likely to develop and whether that is conducive to growth one must understand the dominant ideological features of a society and (iii) that the ideological traits relevant for the birth of private property rights systems with growth are entrepreneurship and self-responsibility. These should not be taken for granted too easily and will now be analysed in greater detail.

The origins of desirable evolutionary patterns

Two key notions identify the origins of success in the fight among competing civilisations. They are the principles of entrepreneurship and of individual responsibility, in which both geography and ideology have played important roles. In particular, growth depends on the willingness and ability of the individual to act, take responsibility for his/her own actions and possibly break new ground (entrepreneurial behaviour).[10] Clearly specified and enforced property rights – private property rights in particular – are of course also necessary. But without entrepreneurship and self-responsibility, property rights *per se* do not generate growth. An ideological or cultural environment hostile to individual responsibility means that individuals are reluctant both to develop new knowledge and to take advantage of their talents, irrespective of the potential for high monetary rewards. Furthermore, such an environment tends to discourage outsiders, who may indeed be willing to take responsibilities, but are afraid that free riders or rent seekers would be morally justified in interfering, if not explicitly encouraged to do so. Stagnation and poverty are the obvious results.

By considering development also as a matter of individual responsibility one is then led to ask why some communities are bound to accept responsibilities more easily than others. Or why some societies just reject any form of responsibility, except for very limited ambits, like the family. This is where

geography and ideology may play a significant role. Of course, there is no such thing as an optimal geographical setting, or an ideal moral standard. They are both highly variable concepts. Individuals move and look for suitable locations to develop and satisfy their aspirations, as the history of migrations shows. Potentially rich lands, such as today's United States or Argentina, have followed entirely different economic paths and generated further puzzles for economic determinism. Similarly, ideologies that may seem beneficial to growth under given circumstances turn out to be obsolete and even counterproductive in contexts where the notion of legitimacy is subject to interpretation and interference by special interest.

Accidents obviously play an important role, too. However, by looking at accidents through neo-institutional lenses one would not go very far, other than describing the different paths that can develop within a civilisation, as shown for instance by De Long and Shleifer (1992). As discussed elsewhere (Colombatto 2003), according to the new institutional school an accident is the starting point of a path-dependent process, whereby the rules of the game develop according to a mix of rent-seeking dynamics and technological events. Although it would be hard to deny this statement, this view ultimately explains evolution by referring to unexplained accidents. That is hardly acceptable as a theory.

By building on the neo-institutional notion of accident, however, modern development economics justifies the introduction of new property rights systems as some kind of favourable event, after which new promising growth patterns can emerge, perhaps thanks to individual successful 'path-finders' that induce imitators to follow (Pejovich 2003). Surely, the assignment and enforcement of property rights and – more generally – the legitimacy of ownership would enhance capital formation and be the clear solution to the development question, as recently suggested, for instance, in de Soto (2000). But the very problem of development is to understand why the legitimacy of private property tends not to be accepted, that is why the hoped-for accident is so unlikely to occur.

Demography and property rights

As regards geography, its effects on growth seem to depend on the response to demographic pressure (North and Thomas 1973) and the need and opportunity to trade (Bauer 1991). Population growth is assumed to have promoted Western growth in different ways. Demographic pressure forced people to migrate towards less-populated areas. Distance made it more difficult for the central authorities to control individuals and thus use violence in order to extract rents. As a result, private property rights were easier to protect against external interference and economic activity was encouraged. Furthermore, individuals living in far-flung areas would be settled in territories featuring

different geological and climatic conditions. This would generate comparative advantages in agricultural productions and profitable trade opportunities, leading to higher living standards. Of course, these effects were considerably enhanced by the discovery of the New World, which also reduced the prospect of famines and malnutrition and therefore allowed further population growth. In turn, a larger population would increase both the potential returns to entrepreneurial innovation (stimulated by a larger market) and the probability that innovation actually takes place (higher number of thinking individuals).[11]

True enough, many of these elements point to the need to remove legal rents as a precondition for growth, which is more or less equivalent to underscoring the need to preserve and protect private property rights. Indeed, rent-seeking situations would dissolve if those who bring to bear the legal monopoly of violence did not infringe upon private property rights.

Nevertheless, this avenue of investigation will not be pursued further. The role of demography is surely not denied. There is no doubt that densely populated areas are more vulnerable to the consequences of bad harvesting and famines and that the scarcity of good land may have been a frequent threat over long periods in the Middle Ages. But these phenomena cannot be seriously taken as a brake on innovation. Indeed, history provides plenty of examples where demography has hardly been a problem for development. On the other hand, it is true that larger markets encourage producers to develop new technologies and reorganise factors so as to satisfy greater demand. Still, although the number of potential beneficiaries has certainly influenced the features of entrepreneurial activities, these numbers hardly had an impact on whether to engage in entrepreneurial activity. Transaction costs (transportation), trade policies and, more generally, the ability to appropriate the return to entrepreneurship were probably far more important.

Similarly, a clearly-defined property rights structure is definitely preferable to a regime of arbitrariness, violence and deceit. Its importance for growth is not denied. However, as already mentioned in previous paragraphs, this does not imply that property rights are a sufficient requirement for economic growth to take place. Nor does it imply that there must be some kind of institutional, top-down intervention so as to define and assign allegedly welfare-maximising rights. Once again, history provides plenty of examples of communities where property rights were fairly clear, where the state played a relatively minor role, and yet economic development was virtually nil.

In a nutshell, searching for appropriate (optimal) property rights structures – including regulation – as recipes for development does not take us very far, and can actually cause considerable harm. More important is the task of identifying the historical periods when property rights became important; of understanding why fairly similar property rights structures and principles for legitimacy led to

growth in some situations, while they remained more or less irrelevant under different contexts. A good part of the answer comes from trade.

Trade

In ancient times, populations that settled in coastal areas and were subject to demographic pressure could expand inland. Or they could trade, either with their neighbours, or with overseas populations. In many cases the first option was viable: coastal populations actually migrated and merged more or less peacefully with inland communities. In other situations sea trade was preferred and contributed to the birth of economic systems based on exchange and specialisation. These economies relied heavily on responsible individuals, who were willing to take risks and were fully legitimised when making significant profits out of their activity.

By contrast, land proximity meant that comparative advantages offered relatively modest rewards to local trade. Exchanges were definitely smaller in quantity and entrepreneurship of less importance. The relatively high cost of land transportation also induced merchants to trade rather precious and easy-to-move goods (gold, silver and slaves, for instance). These features probably conveyed the impression that trade was to the benefit of the élites only (landowners), rather than of the population at large. Hence, entrepreneurs were likely to be less appreciated from a moral viewpoint and the wealth of the merchants had not the same legitimacy as that of the warriors (reward for valour and courage) or – later on – of the aristocracy (honour and, perhaps more questionably, divine will). This difference made envy and expropriation more easily acceptable. If so, it is apparent that private property was bound to be increasingly vulnerable.

In short, before the rise of the so-called 'Western civilisation', which more or less coincided with the downfall of feudal society, entrepreneurship had an opportunity to develop only under rather particular circumstances (sea-trading communities). These could, however, easily break down under foreign aggression and disappear.[12] And even when foreign attacks were repelled, more and more resources (and power) had to be devoted to a central authority responsible for defence. As a consequence, feelings of national, ethnic or even class loyalty became crucial to enhance collective security. When collective security and the absence of domestic tensions mattered, the notion of self-responsibility turned out to be a privilege, rather than a right. It frequently became a prerogative of the recognised rent-seeking classes, that is, of those who were (and are) less prone to develop productive entrepreneurial skills. As a result, entrepreneurship in pre-Christian history was in very short supply (and demand).

When comparing the lack of development in the ancient world[13] with today's undeveloped countries, at least two elements deserve attention. First,

in ancient times entrepreneurship was generally constrained by the difficulties of developing large-scale mercantile societies and by the critical attitude towards self-responsibility – a potential threat to social cohesion and defence capabilities. The role of property rights for development could thus be only marginal. Second, in modern undeveloped countries large-scale entrepreneurship is also discouraged, for it represents a threat to the ruling élites. In particular, trade is allowed only locally, whereas those who want to engage in long-distance transactions are more or less obliged to fall prey to state-run or state-controlled intermediaries. As a consequence, trade fails to develop and producers miss out on all the information that goes with it:[14] about new products, undetected needs, unknown promising markets, new technologies and equipment. Producers who are successful beyond the local boundaries are those who have reached a compromise with the state agencies. The very notion of entrepreneurship is thus undermined, for success tends to be regarded as the result of privilege, rather than of entrepreneurial talent. Once again, the lack of secure property (private) rights is the consequence, rather than the cause.

One might therefore conclude that in ancient societies mercantile economies failed to develop because of the short-term horizon of the rulers, and because of high transaction costs (transportation), which made it difficult to reap the advantages of specialisation. It would also be tempting to add that since today the latter costs are a fraction of what they used to be, the introduction of democracy and the elimination of rent-seeking autocrats would be sufficient conditions to revive entrepreneurship, since it would automatically lead to adequate property rights structures (rule of law). Although presented in slightly different terms, this is actually the core of most current development policies. Unfortunately, this conclusion misses the point and runs counter to the facts.[15]

Politicisation and lack of responsibility rather than property rights

Indeed, the difference between the undeveloped world and the beginning of Western economic growth is not just a matter of lower transportation costs and a huge expansion in the technology frontier attainable. Today's societies also feature a much higher degree of politicisation and democratisation. As forcefully explained in Bauer (1991) and Hoppe (2001), politicisation implies that an increasingly large number of individuals believe that the solution to poverty lies in the political context (better constitutional rules, better policy making, fair property rights rules); rather than in the creation of a suitable entrepreneurial environment that allows agents to strive in order to enhance their own welfare.[16] Politicisation is of course attractive, for it offers ready-made solutions, and diverts attention from the painstaking analysis of the mechanisms that should lead to solutions. A recent example is provided

by Serbia, where most intellectuals believed that the only cause for the country's backwardness was President Slobodan Milošević, so that by removing the tyrant the Serbian economy would blossom with no further action. Instead, the logic of democratic politicisation usually opens up the race to power. Sometimes candidates compete peacefully, especially when the median voter has a lot to lose (high income or wealth) and not much to gain (discretionary power is effectively limited or dispersed). In other cases competition leads to violence, when most of the population has little to lose and much to gain (the winning faction often enjoys virtually unlimited discretionary power). It is then hardly surprising that many countries prefer to avoid democratic rules, and opt for illiberal autocracies, as long as these can limit conflict and violence – even at the expense of growth. Despite common beliefs, the key to development, then, seems to be less politicisation and possibly less democracy, too. Singapore and Hong Kong would of course be fitting examples.

Another element is the role of self-responsibility. In ancient times this was not an issue for a large part of the population. It was more relevant for free men and the élites, but did not lead to significant results, for such results became possible only when self-responsibility was accompanied by entrepreneurship.[17] Prior to that, self-responsibility for the middle-class free man was only applied to his interaction within the local community. The notion of self-responsibility was surely present, since individuals were supposed to behave in the interest of the community and were punished for not doing so. Similarly, they were responsible for complying with the agreements the community had established with other communities. But it is clear that the informal rules of the game in most of those societies were mainly 'defensive', in that they were aiming at consolidating the status quo as a guarantee of social cohesion, rather than at promoting talent and experimentation on a wide scale. Despite this weakness, however, the ancient world actually offered a somewhat more promising picture, compared with today's low-income countries: ancient societies did not dispute the principle of human dignity; they rather disputed the question of who could be defined as 'human'.[18] As a matter of fact, ancient society turned into Western civilisation (that is, growth, from an economic viewpoint) when human dignity became a universal notion and entrepreneurship acquired moral recognition (see also Sirico 2000).

On the other hand, in the last decades politicisation has weakened the notion of self-responsibility, both in the developed and the undeveloped world. And the notion of democracy has favoured the concept of coercive solidarity, which is considered a duty for some, a right for others. Hence, if one takes an optimistic view, one could claim that it will not take centuries to go from backwardness to economic progress: information and transportation costs are low, technological knowledge available is astonishing, and imitation

effects are likely to be sizeable. Still, a more cautious approach would perhaps suggest that today's problem is not about widening and integrating a well-established subjectivist view of social interaction (as shown by the transition from the classical to the Western civilisations). Instead, it is about transforming hopes for utopian, third-way solutions – presumably provided by a new and better class of enlightened politicians – into greater consciousness of the rent-seeking implications of collective decision-making processes. Somewhat paradoxically, the chances of this happening are greater in those areas where the influence of the welfare-state culture is weaker. From such a standpoint, one could actually go further and claim that Westernization is the most dangerous enemy to development, more so than autocratic political institutions or the lack of rule of law.

Ideology

It was suggested above that the development of effective property rights structures depends on the notions of entrepreneurship and self-responsibility. Individuals' attitudes and states of mind heavily influence these concepts. Hence, the notions of economic progress and individual satisfaction cannot be fully understood unless the area of shared ideologies is properly investigated, that is, unless the set of shared moral standards that provide legitimacy to human action is appreciated. One may certainly welcome all attempts to suggest a history of ideologies, which is actually equivalent to suggesting a history of the dynamics of a civilisation. Among others, Oswald Spengler, Arnold Toynbee, John Hicks and Carroll Quigley are celebrated examples. Unfortunately, if one abandons the safe avenues of description, the risk of falling into the traps of social determinism is almost unavoidable.

By contrast, a more fruitful approach consists in looking at the source of ideologies, studying how the legitimacy of individual action is determined, how it evolves, and under what circumstances the existing principles of legitimacy can be replaced by new criteria. From this viewpoint, if a civilisation is identified as a system of shared behavioural rules, then property rights regimes and outcomes reflect both such rules and the environmental features (including geography) with which they interact.

Now, throughout history the sources of rule legitimacy have been either external or internal to humankind.[19] In the former case the norms for social behaviour (morality) come from one or more divinities, who dictate the rules of the game. The power of secular rulers is then legitimate as long as they prove that they have actually been designated by God, that they have not betrayed the implicit pact of obedience to comply with and apply divine rule (call it good government). As we know, this structure of legitimacy provides an effective constraint upon the abuse of power and is therefore a powerful contribution to social stability, as long as divine designation remains unques-

tioned and the pact of obedience is clearly specified and understood by the population.

The other source of legitimacy is humankind as such, as the Enlightenment has asserted. In this case the bond between religion and morality is *de facto* eliminated, and therefore that between religion and power (De Jouvenel [1945] 1993). Individuals themselves define the relevant rules, which may be the outcome of a more or less well-conceived social order (as the constructivist legacy of the French Enlightenment would suggest) or the outcome of spontaneous interactions among free individuals (as the Scottish Enlightenment has claimed). Under both circumstances, however, the ruler is no longer the secular executor of divine will, but either the interpreter and executor of the alleged social contract that underlies the agreed-upon enlightened social design; or the protector of individual freedom.

It has previously been argued that a winning civilisation, that is, a society that survives its inevitable inner tensions[20] and provides opportunities for economic progress, is one where the principle of self-responsibility prevails and productive entrepreneurial talents are rewarded by those who benefit from them. By taking into account the various points suggested so far, the key questions that development economics should then address are thus the following. Can a supposedly optimally-designed society satisfy the two requirements just mentioned? If not, what system of legitimacy does? Where do undeveloped economies currently stand? And what are the chances of them moving from one system to the other? Although it would be instructive, lack of space does not allow a close scrutiny of past historical experiences so as to shed light on the elements that shaped successful dynamics both across and within civilisations. The last part of this chapter, however, will attempt to highlight some elements concerning today's undeveloped economies.

Back to the economics of development – culture or rules?

The search for an optimally-designed property rights structure is indeed the core of state-of-the-art development economics, as taught in the classroom and presumably applied by government and international agencies, including the International Monetary Fund and the World Bank.

Constructivism once again

By extending the discussion proposed above, it is manifest that constructivism *per se* cannot solve the problem of slow economic progress. As a matter of fact, it can make it worse. For the very acceptance of a constructivist approach implies that both the design of the rules and their subsequent interpretations are subject to human discretion, and that the decision about those in charge and about how far their discretion should go is to be defined through the political process. For understandable reasons the mechanism that

is likely to prevail is the one which maximises consensus (democracy). Not surprisingly, the rule-making game is then a mix of rent-seeking interests and redistributive pressures.[21] Such an environment is hardly conducive to growth.

This leads to the essence of the poverty trap, which of course has little or nothing to do with the so-called 'savings' and 'human-capital' gaps. As has been mentioned above, when people with little to lose compete for power, they also have little restraint in transforming peaceful confrontation into violent conflicts. Such conflicts are often described as clashes among different religious or ethnic groups. These explanations certainly contain some truth, but they are only part of the story. The real question remains the acquisition of unrestricted power, an area where self-responsibility, entrepreneurship and concern for property rights enforcement play a very modest role, if at all. Indeed, when it comes to the race to power among competing pressure groups, teamwork and loyalty to the coalition weigh more than responsibility and accountability, and those who better than others succeed in relieving people from individual responsibilities and in lieu promise redistribution through rent-seeking positions (sometimes disguised as 'social rights') acquire group leadership. Furthermore, since the rewards to unproductive entrepreneurs are greater than those obtained by means of productive entrepreneurship, it is hardly surprising that making money by using talent is seldom regarded as an honourable activity. Envy often leads people to associate somebody else's success with unfair behaviour, even when that is not the case.

Under these circumstances private property rights are hardly protected and contracts scarcely enforced. But that account can be no basis for policy, for it amounts to arguing that constructivism has failed because it has not been enforced or because there was not enough of it. Once one accepts the constructivist criterion, the possibility of introducing rule of law in a developing country is seriously compromised. The heart of the matter is that there is no demand for a different property rights approach and most feel justified in adjusting to the prevailing moral environment, including rent-seeking malpractices. The final section of this chapter will address this point.

Legitimacy and development

According to the argument developed so far, the so-called 'poverty trap' of an undeveloped economy is in fact a legitimacy gap, whereby any kind of political leader or élite develops a short-term horizon and feels no restraints in his/her race to power. In the end, in these countries formal democracies and formal autocracies do not differ much. Contrary to what has been argued by some institutional economists (such as Olson 1993), a modern democracy is not necessarily more encompassing than an autocratic regime,[22] for both regimes suffer from the shortcomings of little accountability and from the need to redistribute resources to their clientele.

It is therefore not surprising that public choice scholars as well as institutional authors are somewhat uncomfortable in explaining the birth of an autocracy and then the transition from autocracy to democracy. Indeed, even a cursory look at the literature shows that in order to escape determinism they have to fall back on residual solutions[23] or accidents (exogenous shocks), which is of course hardly satisfactory.

History shows that growth in the West took place when the issue was not about moving from one institutional environment to another, but from one ideological system to another, that is, from one set of sources for legitimacy to another. At the beginning, transition was founded on a unique source of divine legitimacy (the Emperor) being replaced by multiple and often competing representatives of the divine will – the monarch, the aristocracy, the Church. In some cases an oligarchy of learned men also entered the field, thereby advocating the right of the individual to assess the actual meaning of religious obedience, its political consequences and thus enforcing rulers' accountability (Humanism). Contrary, say, to the Islamic world (see, for instance, Labohm 2003), this evolution took place in the West during the Middle Ages. Sometimes it was successful (Northern Italy, the Low Countries), sometimes it almost failed (Spain) or brought about mixed results (France, England).

Later on two other key transitions took place, under the influence of French Enlightenment (and of the French Revolution) and of Romantic Darwinism (leading to the First World War). These caused a reduction in political competition and encouraged constructivism; but they took place at a time when the power of economic progress had already been unleashed, and living standards were already high enough to discourage systematic violent confrontation within each country. More important, the winning coalitions had already established and legitimised themselves as guarantors of enough redistribution to ensure the so-called 'social peace'.

Unfortunately, none of this can be claimed for today's undeveloped world, where the pursuit of rents and policy-making power is assuaged neither by shared moral standards (legitimacy), nor by consolidated coalitions looking for consensus, even if at the expense of a suitable property rights regime and thus growth. Hence, there is no particular reason to believe that the introduction of Western constructivist ideologies will generate much economic progress. It is certainly true that the Western 'rule of law' and property rights enforcement systems are strongly correlated with high incomes. But Western wealth has been created thanks to an ideological environment that is no longer there. Nowadays, much of Western governance structures are designed to stabilise the distribution of that wealth among interest groups and meet both competitive pressures and envy, rather than to enhance the creation of new sources of prosperity. By inducing developing countries to accept such governance structures, Western well-meaning social engineers would induce the undeveloped

world to concentrate on – and fight for – fragments of income, with little concern for the creation of new wealth. Under such circumstances, a heap of desirable constitutions would assign and define more or less questionable rights over very little property.

It thus appears that the best chances for the protection and enforcement of property rights are enjoyed by those countries where an established autocracy draws its legitimacy from its ability to abstain from excessive intervention and to restore dignity to productive entrepreneurship. Only later will this lead to a demand for formal property rights structures and more effective enforcement. Surely, sooner or later, pressure for redistribution will also rise and lead to possible political crises, especially if growth loses momentum. Should pressure come too early, property rights would be violated and economic progress would come to a standstill, if class or ethnic ideologies prevail and large groups have little to lose from violent confrontation. Otherwise, growth will slow down gradually and the catching-up dynamics will come to an end.

Surely, the Western model may also be rejected, as appears to be the case in many European countries, where young people are rapidly losing interest in political engineering and, indeed, in politics as a whole. If so, there are reasons for hope. Although the illusion of the welfare state still has many supporters – even among the young – the wind of globalisation may soon prove too hard to resist. It might encourage individuals to go back to the liberal cornerstones of Middle-Age Western Europe and appreciate the virtues of holding on to the (subjectivist) principles of the natural order.

Notes

* I am grateful to Pierre Perrin, Simon Teitel and to the participants at the Seminar in Austrian Economics at the Faculty of Applied Economics (University of Aix-Marseille III) for their insightful comments on a previous draft of this chapter.

1. Of related interest are the notions of fairness and justice, which differ according to the legitimacy of the action involved. In the Judaeo-Christian tradition, 'just' (that is, legitimate) denotes an action that respects the fundamental principle of equal human dignity, which can be considered as synonymous of individual freedom. On the other hand, fairness reflects the individual evaluation of others' behaviour, which can be approved or disapproved irrespective of its legitimacy. Unfortunately, the distinction between fairness and justice has become blurred in modern democracies, where the notion of 'legitimate' has been replaced by the concept of 'legal' – legal now being what the majority believes to be fair.
2. Many authors, from the seventeenth century (Thomas Hobbes) until today (Buchanan and Tullock 1962) have tried to develop the notion of 'social contract'. All their efforts, however, shared the same purpose, that is, to replace a divine contract with a secular one. As underscored in Hoppe (2001), they all ended up by offering a reason for the monopoly of state power to exercise violence and by justifying the very existence of the state in the absence of divine legitimacy.
3. That includes both physical and non-physical violence (for example, fraud and deception).
4. This is an open issue among subjectivists. The Austrian school (Ludwig von Mises) suggests that one should not try to persuade another individual to accept moral standards other than those spontaneously developed by oneself. By contrast, Economic Personalism

maintains that non-violent calls for moral reform are perfectly legitimate, as long as private property is not questioned. See Beaulier and Prychitko (2001).

5. Despite strong criticism by several Austrian economists (see, for instance, Rothbard 1963 and Batemarco 1987), the literature defines growth as an increase in aggregate purchasing power, following GDP and income statistics, with some adjustments. The term 'development' may take two meanings. It can denote the improvement of a number of aggregate variables that reflect collective well-being, such as per capita GDP, life expectancy, literacy, health conditions, calorific intake, unemployment and infrastructure. Distribution of these variables across the population is also considered, so that homogeneity implicitly turns out to be another important element of development. However, the word 'development' can also denote the range of opportunities open to the individual, where the term 'opportunities' refers to the ability to satisfy preferences and the returns of one's own efforts and investments. The former definition is typical of the orthodox view and is usually described by the so-called 'Social Development Index' (United Nations/World Health Organisation), whereas the latter is typical of the free-market view and is described by the 'Index of Economic Freedom' (Heritage Foundation).
6. The birth of development economics coincides with the downfall of the Western colonial empires and the ascent of the role played by the mass media. In particular, economic events in many former colonies became of international political interest, Western politicians and bureaucrats perceived new opportunities to satisfy their quest for power, honours and prestige, and Western public opinion became more exposed to the many tragedies taking place in previously unknown parts of the world.
7. As recently argued by Carneiro and Heckman (2003), the quality of human capital depends more on the ability to provide motivation and enhance an adequate family environment, rather than on policies to favour education directly.
8. This is generally known as the 'shield effect' (Vaubel 1992), and is also typical of many developed countries, where it often protects bad policy makers from domestic criticism.
9. This is the essence of the so-called 'political economy' approach, as opposed to the institutional school, which emphasises the importance of a bottom-up process. See, for instance, Pejovich (2003).
10. See also Holcombe (2001). Self-responsibility and entrepreneurship are of course two different concepts. The former refers to the full recognition of the consequences of individual behaviour, while the latter refers to the ability to develop new ideas and possibly bear the risk of failure. This holds true when talking about the individual. When analysing the behaviour of a community for an extended period of time, a social environment where entrepreneurship is frowned upon is not likely to encourage self-responsibility. For example, when envy-led behaviour turns out to be morally acceptable, entrepreneurship is likely to be devoted to redistributive (rent-seeking) activities, rather than to the creation of wealth. Thus, if one ignores the entrepreneurial nature of the rent seekers, it is plausible to claim that the ethics of (productive) entrepreneurship and that of self-responsibility go together. This is indeed the view taken in this text, where the notions of entrepreneurship and self-responsibility in a social context are used almost interchangeably.
11. See, for instance, Jones (1999), who also stressed the importance of enforcing intellectual property rights. Indeed, it is not by accident that these were explicitly recognised by legislators in the advanced countries of the time: in Venice as early as the fifteenth century, and in the Low Countries and England about a hundred years later. It should be added, however, that the real target of the legislation to protect intellectual property was not so much free-riding producers, but the Guild system, which *de facto* regulated access to the exploitation and the production of innovation.
12. The idea of an invader with a long-term horizon, and therefore interested in enhancing the wealth of the occupied territories, does not apply to ancient times for two reasons. First, looting and enslavement was the only way for the ruler to remunerate his army and satisfy his clientele. Failure to do so would have been more or less the same as losing power. In addition, the ruler's legitimacy was often questionable. In the best case the ruler was supposed to enjoy the favour of the gods, as witnessed by the various signs to that effect

that the gods were to shower on the population. The incumbent ruler, however, was made vulnerable by the godly signs his competitors might claim to their advantage. Surely, polytheism did not contribute to simplify matters. Christendom eventually solved both problems. God-to-man communication would have one source only (monotheism) and one receiver (or authorised interpreter – the Church).

13. For the purpose of this chapter, 'pre-Christian history' and the 'ancient world' extend approximately until the thirteenth century, when the new religion ceased to affect vast layers of the population and productive entrepreneurship in Western Europe spread with sizeable results.
14. As underlined in Bauer (1991), trade is the engine of entrepreneurship, for it enhances the transfer of knowledge and creates opportunities for new ventures.
15. On the one hand, the correlation between political regimes and growth is tenuous at best. On the other, many autocrats are voted into power through fair elections, and often confirmed to power. In most cases opposition comes from other would-be autocrats, rather than from candidates offering democratic rule.
16. That also includes decisions about investment in equipment and human capital. The presence of allegedly imperfect capital markets is hardly an issue, for local savings are often adequate to finance the investments required. History shows that Western Europe in the Middle Ages started to invest with no support from Venus or Mars. More recently, Barro (1999) has shown that inequality in income distribution has no effect on savings, which is to say that the propensity to save does not depend on income. As regards human capital, Carneiro and Heckman (2003) provide evidence showing that education depends on family ethics and the opportunities for talent to emerge and be rewarded, and much less on government programmes.
17. Of course, the underlying assumption is that self-responsibility is necessary to obtain (productive) entrepreneurship, but not sufficient.
18. As a matter of fact, this side of the question did not disappear very quickly. As late as the sixteenth century the Church failed to take a strong position against those who justified deportation and slavery by claiming that African blacks and South American natives had no soul, and that therefore the rights of human dignity did not apply to them.
19. One area of potential ambiguity must be conceded, though – the case of animism, whereby one cannot claim that people are their own master/mistress except for the rules that discipline their interaction with others. Nor can one claim that these rules come from one or more clearly identified divine authorities.
20. Tensions can be explained by the fact that individuals may share the same broad values, but apply them in different ways, so that temptations to deviate from recognised ethical standards are sometimes hard to resist.
21. Indeed, this problem was not totally overlooked even by some authors of the French Enlightenment, such as Jean Jacques Rousseau, who actually claimed that the people at large are not wise enough to decide, so that a democratic voting procedure would have been inappropriate. The lessons was well taken by the leaders of the French Revolution, who therefore felt authorised to interpret the needs of the people whatever their desires might have been, and actually introduced oligarchic rule first, tyranny later. Especially in the West, twentieth-century politicians chose to ignore that lesson.
22. The very fact that growth seems to be independent of the political regime supports this statement. See, for instance, Gwartney et al. (1999) for a survey of the literature.
23. For instance, Olson (1993) posits that democracy comes to the surface when competing autocrats face stalemate after prolonged fighting and eventually choose a cooperative solution.

References

Barro, R. (1999), 'Inequality, growth, and investment', NBER Working Paper no. 7038, National Bureau of Economic Research, March.

Batemarco, R. (1987), 'GNP, PPR, and the standard of living', *Review of Austrian Economics*, **1** (1), 181–6.

Bauer, Peter (1991), *The Development Frontier*, London: Harvester-Wheatsheaf.
Beaulier, S. and D. Prychitko (2001), 'Does morality hamper the market process? A reappraisal of the Mises thesis', *Journal of Markets and Morality*, **4** (1), Spring, 43–54.
Buchanan, James and Gordon Tullock (1962), *The Calculus of Consent*, Ann Arbor, MI: University of Michigan Press.
Carneiro, P. and J. Heckman (2003), 'Human capital policy', NBER Working Paper no. 9495, National Bureau of Economic Research, February.
Colombatto, E. (2003), 'Towards a non-Darwinian theory of institutional change', *Journal of Bioeconomics*, **5** (1), 1–25.
De Jouvenel, Bertrand (1945), *Du Pouvoir: Histoire Naturelle de sa croissance* [On Power: The Natural History of its Growth], Geneva: Les éditions du cheval ailé (reprinted in English in Indianapolis, IN: Liberty Fund, 1993).
De Long, B. and R. Shleifer (1992), 'Princes and merchants: European city growth before the industrial revolution', NBER Working Paper no. 4274, National Bureau of Economic Research, December.
de Soto, Hernando (2000), *The Mystery of Capital*, London: Bantam Press.
Gwartney, J., R. Lawson and R. Holcombe (1999), 'Economic freedom and the environment for economic growth', *Journal of Institutional and Theoretical Economics*, **155**, 643–63.
Holcombe, Randall (2001), 'The invisible hand and economic progress', in Gary Libecap (ed.), *Entrepreneurial Inputs and Outputs*, Amsterdam: Elsevier Science, pp. 281–326.
Hoppe, Hans-Hermann (2001), *Democracy – The God that Failed*, New Brunswick, NJ: Transaction Publishers.
Jones, Ch. (1999), 'Was an industrial revolution inevitable? Economic growth over the very long run', NBER Working Paper no. 7375, National Bureau of Economic Research, October.
Labohm, H. (2003), 'Islam and development', Techcentralstation Europe, 18 February http://www.techcentralstation.be/2051/wrapper.jsp?PID=2051-100&CID=2051-021203M, 21 February.
North, Douglass and Robert Thomas (1973), *The Rise of the Western World*, Cambridge: Cambridge University Press.
Olson, Mancur (1993), 'Dictatorship, democracy and development', *American Political Science Review*, **87** (3), September, 567–76.
Pejovich, S. (2003), 'Understanding the transaction costs of transition: it's the culture, stupid', *Review of Austrian Economics*, **16** (4), December, 347–61.
Rothbard, Murray (1963), *America's Great Depression*, Kansas City, KS: Sheed & Ward.
Sirico, Robert (2000), 'The entrepreneurial vocation', *Journal of Markets and Morality*, **3** (1), Spring, 1–21.
Vaubel, R. (1992), 'The political economy of centralisation and the European Community', *Journal des Economistes et des Etudes Humaines*, **3** (1), March, 11–48.

12 Germline engineering: whose right?

*Lloyd Cohen**

Introduction

The issue to which I shall direct my attention in this chapter is human 'germline engineering'. This term refers to the coming prospect of changing the genetic code of human embryos. There are many within the community of philosophers, scientists, 'medical ethicists' and laymen who oppose such tinkering. I do not; I relish the prospect. If there is some justification in prohibiting germline engineering it must rest on the evil or harm of the activity. I make a distinction between evil and harm to allow for the possibility that some action might fall in the first category (evil) while not falling in the second (harm), that is, that there is a moral dimension to human action independent of its effect – or intended effect – on other human beings. But because as a lawyer and economist I can offer no special insight into evil, I shall restrict my inquiry to harm.

My ultimate argument flows from the outcome of the analysis of the various potential harms alleged to flow from germline engineering. While I find that there may be minor categories that are problematic, these are no more than quirky – almost bizarre – exceptions to the central case, that is, that germline intervention will yield an enormous improvement in the human lot. Those exceptional instances in which one could imagine germline intervention resulting in harm to the human condition or prospect neither require nor justify any significant restraint or prohibition of germline intervention in general.

Some readers may agree in principle and yet still wonder whether at least regulation that might lightly brake the headlong rush of germline engineering might be prudent. Should we not proceed at a more judicious pace? Consider the following observation about casualties in war. After a successful battle the official report states that casualties were light. For those who were killed, however, casualties were 100 per cent. So too with germline intervention; for those who could be relieved of hideous suffering and early death by germline intervention delay means that they pay an almost unimaginable price, and more importantly a price that I shall argue is not justified by any danger that is posed.

Genes and property

The issue I address is the right to manipulate human genes. Genes, human genes in particular, exist only as expressed in actual living organisms. Those organisms and the genes they possess and express are a form of property. To describe something as property is to imply a set of (property) rights that adhere to it. The question of the right to manipulate human genes is ultimately a question of property rights. Who or what have (or should have) rights with respect to human genes embodied and expressed in individual people? Let us begin by clarifying the scope and nature of property rights in general, and those in human beings in particular.

Property rights in human beings

Property is every 'thing' that people care to have rights with respect to. It is useful to think of the different sets of property rights that can and do exist as bounded by a triad, in which any particular property right must be some variation on a theme defined by the three endpoints of the triad. Those endpoints are: (i) private property – property with respect to which a single person has the right to exclude, use and alienate, for example, my apple; (ii) communal property – property which everyone has an identical right to use and from which none has the right to exclude another or a power to alienate from himself, for example, the air we breathe; and (iii) collective property – property which some political body has the right to alienate, exclude and define the set of permitted uses and terms of access; these uses and terms may be as limited and quasi-private as those of the space shuttle *Atlantis* or as broad and quasi-communal as state forests.

A property right need not be purely communal, or perfectly private, or wholly collective. It may partake of some of each category. But, out of logical necessity, it cannot extend beyond the boundary of the triad. That is, rights to exclude must be either private, or collective, or non-existent; permitted uses must be determined either collectively, privately (subject to collective constraints), or not at all; and rights to alienate must be either private, or collective, or non-existent. Every legal system and its defenders recognize some of each category of property. They differ merely as to proportion and detail, which is more than enough to drive men to the barricades.

What of the human body? Where does it lie in the triad? In the modern world, probably the single most universal, uncontroversial, and passionately held right recognized by all civilized people is that each of us owns our selves. In our time and place, where slavery has come to be seen as a moral outrage, each will expressing itself through a particular body is deemed to be the person who owns that body. Thus, for example, a medical researcher must obtain the consent of the subject because it is the subject's body that is to be

investigated, and the subject has one singular unambiguous right with respect to that body, the right to exclude others from all but trivial contact with it.

Consider the following thought experiment. Imagine that a large number of men (or alternatively women) were marooned with but a single woman (or alternatively man). Some might argue for a moral imperative to equal access to that singular sexual and procreative resource. I offer this scenario to show how extreme and far-fetched the example must be in order to generate even a hint of plausibility to the notion that a person's body could possibly be a communal asset. The extreme character of the example highlights how very strong is our moral intuition that if anything in this world sits at the vertex of private property it is the human body and that each person owns their own.

Individual threats to the collective

Although the liberty and autonomy of the individual with respect to his own body sits at the center of our modern enlightened system of ethics, reasonable men recognize some limited collective rights with respect to the bodies of those who make up the collective. Consider incarceration, quarantine, conscription, and police searches and seizures of the body including even the bodily fluids of suspects. Each of these can be considered a collective right that the group takes in the body of some of its individual members. Each involves a serious breach of the liberty and autonomy interests of the individual by the collective.

The example most apposite to restrictions on genetic engineering is medical quarantine. A person is quarantined when organisms within his body pose a serious physical threat to the well-being of other members of the community. Quarantine, like all collective impositions on the private property rights that an individual has in his own body is an exception triggered by exceptional circumstances. Note also that quarantine, like the other impositions by the collective on the individual are not the privilege of every collective to which the individual can be ascribed, but rather of only one class of collective, the sovereign authority. Churches, ethnic groups, business associations, bowling leagues, linguistic groups, country clubs, and all the other varied collective entities that one might find oneself a part of are not recognized as having a right to incarcerate, quarantine, or conscript. At most they may exclude.

The right of the sovereign to quarantine is a constraint on the primary right that is lodged in the individual. The constraint arises because the unconstrained action of the individual will constitute a substantial threat to the welfare of the group. Placing restraints on the individual's use of his body is in this respect a mere variation on the more general theme of restrictions on the use by the individual of other property over which he has dominion. Just as our right to own explosives and toxic materials does not mean that we are

free to store or move them in any manner we choose, so too states and communities may quarantine people if they have a reasonable fear that the individual may be infected with a contagious disease. Because we each pose dangers to one another through the use of property other than our bodies, such limitations on individual rights over one's property are so common that they are usually taken for granted, and not even recognized as a collective impingement on private property rights.

Germline engineering

Germline engineering is the process of 'artificially' and purposefully changing the genes of an organism such that when it procreates the changes replicate themselves in the next generation and – subject to the laws of genetics – in all future generations. The prospect of germline engineering has been fiercely criticized. Much of this criticism is directed at the 'un-natural' nature of this form of intervention. As an economist and a lawyer I am unequipped to speak on the question of whether this process would be 'natural', and whether or not the answer to that question has any moral, and *pari passu* legal, weight. Instead I shall address the question of harm and benefit.

Categories of harm

Three categories of harms have been ascribed to germline engineering. The first is a fear that prospective parents will simply make poor choices that harm the genetic inheritance of countless generations of their issue, that they will produce a phenotype that from the beginning is – or should be – seen as inferior to what would have otherwise naturally occurred. The second is some concern with inequalities (and therefore inequities?) in access to genetic advantages, that is, that germline engineering will not be available to all and so will allow some to provide their offspring with an 'unfair' advantage over others; and the third is a general concern with the integrity of 'the gene pool', that is, that the variety of human genes will be reduced, thereby creating some prospective danger to human beings as a species. We shall examine each of these categories in turn by considering several hypothetical instances of germline engineering that might give rise to them.

Engineered monsters

The first category consists of parents intentionally choosing to genetically transform their progeny in a manner that the wider community considers harmful to the progeny. Consider the following:

- *Case 1*: In China in the not too distant past mothers bound the feet of their infant daughters to stunt their growth. It was considered most attractive for a young woman to have a foot small enough to fit into the

proverbial teacup. In addition it was a way for the Chinese to assert their national pride and to distinguish themselves from, and refuse to integrate with, their Mongol Conquerors. Imagine that some parents chose to genetically transform their female embryos to yield girls with tiny feet.[1]

- *Case 2*: Female circumcision (clitorectomy) is practiced in some Arab communities, with the effect, and for the apparent purpose of reducing the sexual satisfaction of women. Parents concerned that their daughters maintain their chastity might alternatively seek to genetically obstruct the development of the clitoris.
- *Case 3*: Some writers have speculated that deaf parents might choose to have deaf children (Heller 2001, p. 165).

Note how almost frivolous and fanciful these illustrations appear. I did not try to make them so; it is simply very difficult to conceive of serious examples that satisfy this criterion. Why should that be so?

Mistreatment of children arises from two sources; bad motives, and bad judgment. The generic bad motive is selfishness; some parents care too little for their children and neglect them, and others use their children for their own narrowly selfish purposes and thereby abuse them.

Neglect finds no expression in germline engineering. If prospective parents do harm their offspring through germline engineering it is from the opposite of neglect, unwarranted meddling. But what of the analog to abuse? There is no analytic a priori reason why abuse cannot find an expression in germline engineering. Just as parents might work their children to death, or rent them out to others, so too they might genetically engineer their germline for their own selfish benefit rather than that of their offspring. But even if evil prospective parents were to choose this path, it is more than a little difficult to conceive of how a third party would detect it. Even – perhaps especially – exploitative parents would choose genes for their children that make them as healthy, strong, vigorous and intelligent as possible. Exploitative parents would want their child to be of the highest value – all the more to exploit![2] As long as the asset value of the child is maximized in the same markets where people operate to maximize their own value, the genetic choices of egregiously selfish prospective parents will be indistinguishable from those of selfless parents.[3]

So disposing of bad motive, we are left with bad judgment; that is, prospective parents mistakenly making bad genetic choices for their offspring. What sorts of 'mistakes' might they make? We can gain some insight into this by first looking at the post-birth 'mistakes' that parents make. Some examples that come to mind are: (i) members of some religious groups refusing medical treatment for their children; or (ii) not allowing their children to be

educated beyond the 6th grade; or (iii) as previously mentioned, female circumcision. What strikes one immediately is that such cases are far less common and public intervention is far more morally and politically problematic than cases of abuse and neglect motivated by selfishness. Why are they so rare, and problematic? And how does that inform the issue of germline manipulation?

Their rarity rests on our common humanity. Because of our genetic predispositions and our shared culture, if we care at all for our offspring, we have broadly similar goals for them. The reason that public intervention is so problematic even in those few cases where the goals of the parents are so at variance with the norm, is that respect for the prerogatives of parenthood is a central cross-cultural value. So what does this imply for germline engineering?

First, egregiously bad genetic choices by prospective parents are likely to be trivial in number. As human beings, while we may differ markedly in some cultural characteristics, such as religion and language, we have evolved a set of values and tastes that are broadly similar; we all value health, we all value intelligence, and we have similar, albeit not identical, aesthetic tastes. To be more concrete and graphic, unlike many avian species our infants do not routinely engage in siblicide, and unlike a number of small mammals, human mothers do not routinely eat their young. Put simply, all human beings share a largely common human nature.

Second, because the moral range of germline engineering is less than that of the later socialization and upbringing of children, the number and variety of potentially offensive germline modifications will be still more limited; one can educate one's children in forbidden political philosophies and religions but one cannot engineer their genes to the same effect. Thus while the thesis that parents will seek to manipulate their germline so as to have children who are stupid and unhealthy violates no analytic a priori proposition, it does stretch the limits of plausibility.

That said, let us accept for the sake of argument that we are not dealing with an empty set, that is, that some prospective parents might choose to engineer their germline so as to have children with genetic characteristics that the wider community believes do not serve the offspring's interest, for example, pointed heads, tiny feet, no clitoris, deafness. What then? Does the prospect of this danger require some supplementary regulation or prohibition?

Who is the proper judge of the interest of the as yet unborn child? The answer I offer without further argument in support is that it is the child herself. And there is the rub. The child is not there to express her considered judgment. Germline engineering is, however, not unique in presenting this difficult obstacle. Infants, small children, the mentally retarded, disturbed and disabled are in much the same position as embryos. They similarly are not in a position to inform us of their preferences, to say nothing of a

considered judgment of their interests. So whether a mother engineers the genes of the embryo that will become her daughter so that she is born with tiny feet, or alternatively binds her daughter's feet after birth, in both cases the embryo and child are too immature to express an opinion on the question. Someone else must offer the judgment and make the decision for the child or future child.

As a general matter, who better than the parents to judge and represent the best interests of their prospective issue? We entrust parents with the power to make all manner of decisions that affect the future health, character and personality of their children. What justifies this deference? Three bases suggest themselves. First, of all the adults who might exercise judgment on behalf of the child it is the parents who are most likely to have the child's interest at heart. Second, those who assume the duties of parenthood are entitled to some compensating privilege. And third, parents have a better sense of what is good for the child than the child does (Dworkin 2001, p. 153).

I see nothing that warrants a more intrusive policy for germline engineering. Indeed, there are three substantial reasons to allow greater deference to prospective parents in germline engineering of their embryos than to actual parents in post-birth choices for their children. First, there are fewer and less plausible base conflicts of interest (financial in particular) at play. Parents that refuse to expend resources on the education or health care of their extant child might claim to do so for religious reasons, but it also serves the parents' own financial interests. A corresponding niggardly bent in the germline case would only result in declining to engage in germline engineering, rather than some disapproved choice of genotype.

Second, the earlier in the chain of events leading to the realization of personhood that one acts to change the outcome, the less the ultimate person affected has a claim against the actor as a matter of right. How else justify abortion? And why else distinguish late- from early-term abortion? The germline engineering decision occurs at the earliest stage of the development of the human being when its ability and right to decide for itself is as limited as it could be, and furthermore the decision cannot be postponed.

Parents do not own their children. Human beings are not the property of anyone, even their parents. But parents are granted broad rights to the care and upbringing of their children. As the child matures the prerogatives of the parents diminish and the rights of the child increase. By extension back, the 'child' that is but a mere embryo, or a member of the even more inchoate future generations of offspring, is entitled to even less deference as against the judgment of the prospective parents. Thus in comparison to either male or female circumcision, genetic engineering to achieve the same result seems less morally problematic as a usurpation of the rights of another human being.

Third, if there are such things as fundamental rights that deserve heightened protection from state interference, and there is a hierarchy of those rights, then while child-rearing decisions fall well within the class, procreation must lie even closer to the center and deserve even greater deference. Procreation is a central primal act of not only human beings, not only primates, not only mammals, but of all living species. As such, the right to procreate or not is granted the broadest protection by, and deference from the state, for example, constitutional rights to marry whom one chooses, acquire contraceptives, and abort pregnancies. Even those who carry known genetic defects are perfectly free not only to engage in sexual intercourse but to procreate without interference from the state. Whatever deference is appropriate to parental choices for their extant child with respect to circumcision (male or female), education, or religious training, an even greater deference is due to achieving those or other ends through procreative decisions.

So I conclude that the reason we can think of only the most bizarre and fanciful examples is that this category of potential harms is virtually a null set, and as for those few cases that might arise our best policy is tolerance and acceptance.

Inequitable advantages

If the first objection to germline engineering is the concern that prospective parents will not do a good enough job representing their offspring's interests, the second objection is its opposite, that they will do too good a job, that is, that they will provide their own offspring with a genetic advantage not available to others. To those morbidly fixated on the issue of equality this may seem a substantial problem.

Evaluating the potential harm of genetic manipulations that provide a benefit to some not made available to others requires a taxonomy that allows one to place each illustrative example into its own meaningful class. The first distinction that will have power to some readers is that between an equalizing transformation and an unequalizing one. Eliminating an inherited genetic disease would be equalizing, while fostering an immunity to colds is potentially unequalizing. The vast majority of human beings do not carry the gene for Huntington's disease. A germline intervention that eliminates this gene from those who carry it would be equalizing in that it would improve the lot of the minority and place them on a par with the majority. There can be no egalitarian objections to genetic manipulation that benefits those worse off while causing no harm to those better off.

But what of unequalizing germline interventions? Imagine that it is possible to provide a genetic immunity to the viruses that cause the common cold. Is it acceptable for a minority of potential parents to purchase this for their offspring, or, alternatively, for the state to provide this to a minority of

embryos? Whether we consider colds, intelligence, stature, strength, beauty or a host of other advantages that could potentially be fostered by germline manipulation, if only a minority of the population has access, a more dispersed distribution of positive human characteristics will result. Those who place a high premium on equality may find this objectionable.

But all unequalizing improvements (or for that matter equalizing ones) are not the same. They can further be classified by how the change in phenotype tangibly affects others not themselves subject to the genetic change. There are four possible economic classes arrayed along a scale moving from: (i) Pareto improving, through (ii) social wealth increasing, to (iii) social wealth neutral, and ending with (iv) social wealth reducing.

A Pareto-improving genetic transformation will make the subject better off while making no one else worse off, and more likely making them better off as well, though almost certainly yielding them less gain than that of the actual subjects of the genetic transformation. Examples of Pareto improvements from other spheres of life are legion. Consider the discovery of the polio vaccine. It brought fame and fortune to its discoverer Jonas Salk while at the same time protecting millions of children from a horrific disease. Everyone gained but some – especially Salk himself – gained more than others.

The second class is the social wealth-increasing transformation. Some win, others lose, and the value of the gain to the winners (measured in monetary units) is greater than the loss to the losers. This class, along with the first (Pareto improving), are the engines of human progress. Consider the invention of the hand-held calculator. It improved the welfare of all who made it, all who used it, and a legion of others who transacted with them. But not everyone gained. The advent of the calculator came at the cost of the livelihoods of those who had produced and skillfully used sliderules.

The third class consists of social wealth-neutral changes. They benefit some but at an exactly (or nearly so) equal loss to others. Consider the effect of someone arriving early to purchase underpriced tickets for a concert or sporting event. Given that there are only a finite number of tickets on sale, the earlybird's gain is exactly the latecomer's loss.

The final class consists of changes that benefit some but at a greater cost to the rest of the population. Consider steroid use by competitive athletes. Assume that the goal is not an improved athletic performance *per se*, but defeating one's rivals. In some sense this is much like the previous example; my gain from the use of steroids (victory) is exactly your loss. But the health price paid by athletes who ingest steroids is high. Thus the net gain to those who indulge is less than the loss to those who abstain. Were all to take steroids and not change their relative performances all would lose from the availability and use of steroids.

While some will object on egalitarian grounds to all unequalizing germline manipulations regardless of which economic class they occupy, and others will object to none, most observers will favor those that appear in the earlier classes over those that appear in the later ones. So let us examine some illustrative cases of genetic modifications and consider which class they rightly occupy:

- *case 4*: fostering immunity to viruses that cause the common cold;
- *case 5*: increasing intelligence;
- *case 6*: increasing physical beauty in females;
- *case 7*: increasing aggression and selfishness in males.

These four examples are meant to imprecisely represent candidate cases respectively for each of the four classes discussed above – Pareto improving, social wealth increasing, social wealth neutral, and social wealth reducing. But as we shall see, arguments are available – indeed I would, and will, make those arguments myself – in favor of placing each case in a different class. I am less concerned about persuading you of the proper placement of each case, than in illustrating the underlying principle that a genetic transformation could conceivably fall into each of the four classes, and in suggesting that widely held views on prohibition and regulation will turn on where a candidate germline intervention is believed to properly fit.

Let us begin with case 4, fostering immunity to the common cold. Such genetic immunity would not only be a great benefit to those whose genes are transformed, but would also profit those whose genes remain unchanged. It would reduce the quantity of viral matter spread by those around them. This is similar to the effect on one's likelihood of falling victim to polio if everyone else in one's community has been vaccinated against the disease; there are fewer hosts for the virus, and thus even those not immunized are far less likely to contract the disease. Returning to the distinction I made earlier, unless one believes that increased inequality is *per se* 'evil', even when it benefits all, I can conceive of no reason why anyone would object to a germline change that created immunity to the common cold.

Case 5 is germline intervention to increase intelligence. Intelligence to my way of thinking is a good unto itself, to be valued in its own right. But even if it is not, all the other forms of human wealth – all that humans create and then value – are built on the back of human intelligence. Were it not for our intelligence we would still be living like chimpanzees. It is beyond contravention that the more intelligent the population the larger the economic pie. It is likewise indisputable that as a general matter the greater one's intelligence, the proportionally larger slice of the pie one can acquire (Herrnstein and Murray 1994).

The more empirically difficult and perhaps ethically important question will be whether making one subgroup of the population more intelligent increases the size of 'everyone's' slice. I place 'everyone' in quotation marks to capture the ambiguity as to how many others, and which ones, need be hurt before we think we have crossed the fuzzy line between a merely social wealth-increasing change and one that satisfies some less than absolute sense of Pareto improving – a standard that suffices as a matter of practical political and moral philosophy. That someone would not receive a chaired professorship in physics at the Massachusetts Institute of Technology (MIT), and must settle instead for some slightly less grand position because there are now a spate of brilliant physicists generated by germline engineering, as a practical matter surely does not condemn us to the morally problematic universe of the social wealth increasing rather than the more pristine and exalted regions of the Pareto improving. In my view increasing human intelligence – even of a minority – would be the immensely powerful rising tide that would raise all ships with the possible and limited exception of the apocryphal un-chaired physics professor, and so is an unambiguous *ex ante* benefit to all.

I would go further and say that as a political and moral matter, probably little rests on the distinction between Pareto-improving and social wealth-increasing forms of germline engineering. Does anyone have a right to their place in the genetic hierarchy? If not, then just as you would have no claim to seek an injunction to prevent me from providing my children with a superior education or better nutrition, so too you should have no claim to prevent me from providing my children with more intelligence via superior genes.

Further, there is a common moral and political distinction, embodied in, and exemplified by, the jurisprudence of the 'Takings Clause' of the US Constitution, between property, the taking of which entitles us to compensation, versus some mere expectation of future wealth if the world remains as it is for which no compensation is warranted. The physics chair at MIT that you would have received but for the genetic enhancement of your rivals falls into the latter class rather than the former. So not only do those who would lose from the positional characteristic of this form of genetic enhancement of others have no right to prevent it, they are not even entitled to compensation.

As a general matter social wealth-increasing changes, even if not strictly Pareto superior, should be strongly favored. Were this not the practice of our forefathers our lives would be primitive indeed. The great history of human advancement rests on the many millions of small social wealth-increasing innovations that individually yielded gains to the winners only slightly larger than the losses to the losers; in the aggregate, however, they have yielded an enormous gain to all later generations. Consider, for example, each iteration of the ballpoint pen that drove prior manufacturers out of business and left unemployed resources in their wake. This Schumpeterian 'creative destruc-

tion' resulted to this point in the transformation of a messy extraordinarily expensive ($25 in 1945 dollars) novelty into an efficient, reliable and inexpensive writing instrument.

Case 6 is genetically engineering more beautiful female offspring. I have suggested that this falls into the social wealth-neutral class. That is, it would function much like arriving early to purchase tickets for a popular concert thereby precluding others from purchasing. The analogy rests on a notion that female beauty is essentially a currency employed in a zero-sum game. Women compete for a fixed amount of attention (dates, marriage and so on) from men, and one woman's gain in this game is another's loss.

Undoubtedly, some readers will be offended by this example. Ironically, the illustrative power of the example is intimately tied to its contentious character. The propriety of using germline intervention to enhance the aesthetic appeal of women will turn for each of us on how we see the function of such appeal.

Plausible arguments can be given that such a germline transformation belongs more properly in the class below or the one above social wealth neutrality, or even in the earlier grand category of parental error. Anyone persuaded of the applicability of the model of competition for a fixed resource – men's attention – might ask why this is not a negative- rather than a zero-sum game. After all, we have presumed that such germline engineering is not available to all because it is costly. That cost must enter the utility calculus, and in doing so it transforms this competition into a substantially negative-sum game. On the other hand, others might more fundamentally object to the assumption of an exclusively, or at least overwhelmingly, 'sexist' and sexual motivation for this genetic manipulation. Is not beauty a good in itself to be treasured and valued for itself? If so then such genetic enhancement will be social wealth increasing. And finally, some will object to anyone participating in the female attractiveness competition on the grounds that it reinforces a pernicious view of women.

To repeat, the propriety of each particular instance of germline engineering will, for some, explicitly, and for many others, implicitly, turn on which category and class it properly occupies. But as this example reveals, that placement in turn rests on the ontology of the judge, and there are abundant arguments – sensible and tendentious – available to cast a candidate in either a suspect or protected light.

Finally we have case 7, genetically engineering more aggressive and selfish men. In this case as in the one above the genetic transformation is ultimately motivated by a virtually universal 'desire' of all living organisms to spread their genes. Men who are more aggressive and selfish will for a variety of reasons generally be able to gain access to more women for procreation. In the simplest archetypical case they may simply violently displace a rival.

Such increased aggression would lead to a more Hobbesian world, one that even Leviathan would find most difficult to productively manage. Each prospective parent's gain in creating an aggressive son imposes a significantly greater cost on the rest of the community. Here too one can make arguments that this depiction is too dismal, that an aggressive nature is a positive characteristic in enterprise and science. It is less important whether you agree that this particular case is properly placed in the bottom class, than that you are persuaded that it is at least a plausible candidate in the two senses that prospective parents might well choose such a transformation and it very well might be social wealth reducing. If so then we can agree that social wealth-reducing germline interventions is not an a priori null set.

It seems to me that this last class, and perhaps the one above, are plausible candidates for not merely regulation, but even prohibition. Athletic governing bodies prohibit the use of anabolic steroids which have the effect of providing a competitive boost to some at the equal loss to others, all coming at a substantial price in health to the user. More generally, societies employ a variety of legal and social institutions – tort law, criminal law, manners – designed to limit social wealth-reducing activity. In that same spirit, so too might it ban social wealth-reducing germline intervention.

Such a prohibition would in no way rest on germline intervention being unequally available. Indeed, the case for prohibition is stronger if it is equally available, for then a ban on its use would be Pareto improving – all would gain!

Now let us look at these four cases and the four classes in which we have placed them as a whole. I think it not accidental that the examples in the first two classes seem much more plausible choices and scientific possibilities than those in the last two. More important, the former seem to be mere samples of a much broader universe of potential genetic manipulations (for example, after genetically creating an immunity to the common cold, we might add AIDS, and cancer) while the latter are quirky, tortured, *sui generis* special cases. The general point is that whatever policy we might have to deal with the social wealth-reducing cases, I think we may rest assured that those potential germline interventions are a tiny proportion of the likely candidates.

There are two further observations I offer with regard to privately beneficial genetic changes: (i) in practice they are not likely to be as inegalitarian as some might fear; and (ii) we already permit private decisions that yield far greater predictable equality-reducing genetic outcomes. Germline manipulation will in the end be more egalitarian than it now appears for two reasons. First, and most important, like all technologies average costs will decline with volume and learning. Just as modern technology has reduced the cost of food, ballpoint pens, the printed word, computers and calculators, so too it

will reduce the cost of germline engineering. And therefore just as books, which were once exclusively available to the super-rich, now may be purchased by all, so too germline engineering will before long become available to the mass of men. Second, germline engineering essentially consists of replacing protein bases with known successful alternatives. Those alternatives will be known because of the phenotypes they produce. Thus those with the most successful genetic structures will have the least to gain from germline engineering. Indeed, it will be their genes which will be copied.

The second point is that if some are truly troubled by the possibility of increasing genetic inequality generated by genetic engineering, they should turn their attention to a far more powerful source of increasing genetic inequality, assortive mating. We are free to pick our own mates. Higher status men and women (however status is measured for each sex) tend to choose one another. Procreation through sexual intercourse allows us to not merely add a few favored alleles to the genetic mix of our offspring, but instead to combine an entire set of perhaps 34 000[4] genes with our own. And, we do this not at some substantial cost but as an ancillary benefit to the otherwise pleasurable act of lovemaking.

My brilliant friend David Friedman is the son of the Nobel prize-winning economist Milton Friedman and his wife Rose (Director) Friedman, a woman of sterling intellectual ability in her own right.[5] David's intellectual brilliance is a direct result of the mating choices of his parents. Would anyone suggest that superior people not be permitted to procreate with one another? Such permissible voluntary activity does infinitely more to further genetic inequality in future generations than any conceivable germline manipulation. Germline manipulation will at most add trivially to this inequality. More likely, however, as I suggested above, given the expected rapid decline in the cost of genetic therapy over time, the ability to enhance the genetic virtues of one's offspring will become widely and cheaply available and thus serve to equalize our genetic endowment.

So I conclude this section by noting that it is precisely the improvement of the genetic makeup of people that is the great hope of germline engineering. Potential inequalities that might be exacerbated are: (i) trivial in number; (ii) likely to nonetheless improve the welfare of those at the bottom of the distribution; (iii) expected to increase the wealth and well-being of humanity as a whole; and (iv) swamped by the inequalities already generated by the marriage and procreative decisions of free people.

Endangering the gene pool

The final, and most frightening objection to germline engineering is that it will result in some nightmarish metamorphosis of 'the human gene pool'. I believe, to the contrary, that as with a literal nightmare, the danger is illusory.

The 'human gene pool' is nothing more than a wonderfully evocative metaphor for a rather mundane concept, that is, the number of carriers of human genes capable of procreating and the variation in those genes across that population. Thus the size and character of the gene pool is measured along at least three dimensions: (i) the number of fertile gene carriers; (ii) the variation in the genes they carry; and (iii) the relative frequencies of the carriers of those variations.

The current size of the pool and the particular distribution of genes within the pool are neither durable nor sacrosanct. The human gene pool has been changing since time immemorial and will continue to do so in the future. The number of fertile carriers is essentially a linear function of the size of the human population, increasing in parallel over time, most dramatically in the last 200 years. The varieties of genes and their relative representation in the population have gone through a more ponderous and fundamental transformation. The very metamorphosis of a species of upright apes into human beings was nothing more than a shift of the central tendencies of a gene pool through natural and sexual selection.

Were there a virtue in preserving the current gene pool there would be no hope of doing so. It will change in the future – for good or ill – driven by the same forces that have brought it to its current state. Germline intervention, because of its limited practice in the foreseeable future, can have no more than a minor marginal effect on the pace and direction of change of the human gene pool.

But, looking unimaginably far into the human future and assuming for the sake of argument that germline engineering could have a major effect on the gene pool of the entire human race, of what consequence would those changes in the gene pool have for any one of us or our progeny? The impact of the gene pool on the welfare of any human being arises from two related sources. The principal one is a micro-concern. The gene pool presents each of us with better or worse prospects for a successful outcome to our efforts to procreate. More concretely, I am concerned that my two daughters and son find suitable mates, where suitability entails among other things some genetic characteristics that promise good health, intellectual achievement, physical vigor and aesthetic appeal for their offspring. The gene pool sets the limits and likelihood of finding such a mate.

If much of the rest of the human race were to make systematically bad germline choices then my children would have fewer potential suitable mates. But, it is hard to imagine that anyone – let alone everyone – will engineer germlines that result in changes of genotypes and phenotypes that I found so unappealing that I would reject those people as mates for my progeny.

But assuming the bizarre and unlikely, that others did make choices I found perverse, what then? The people and peoples of the world already

differ in their genes and gene pools, respectively, to a far greater degree than could (in the foreseeable future) be brought about by genetic engineering. Does that significantly affect the procreation prospects of my children? Procreation is not a random process. Many people carry genes for bad health, stupidity and homeliness. Those people, generally, are themselves unhealthy, stupid and homely, and so will be unattractive both as life mates and for their genetic prospects. But they are no threat to us, for none of us need mate with them. So, even if others made poor germline choices, the rest of us are forewarned and may look elsewhere for mates.

Perhaps it is the second – macro – source of danger to the gene pool that we should fear. The ostensible danger is that a reduction in human genetic variety may leave us vulnerable to a cataclysmic health disaster. We are social animals. Germline changes that threaten the very survival of the human race would leave even the survivors in perilous circumstances. Advanced human civilization rests on specialization and exchange. A human society consisting of but 50 000 people, instead of the current 5 000 000 000 would revert to a primitive economic state.

But why should germline engineering lead to such a cataclysm? The notion is that a move toward 'superior' genes may, *pari passu*, mean a move to genetic uniformity, and in the end threaten the robustness of the human race as an entity. Microorganisms (bacteria, viruses and fungi) are constantly attacking human beings. We evolve immunities and the microorganisms evolve to undermine them. The great variety of human beings makes it more likely that there will always be a large saving remnant that by chance has a gene-based immunity to AIDS or some other plague. Human genetic variation ensures that that remnant will be of sufficient size to preserve not only life, but civilization. The fear is that if germline transformation becomes universal, the tendency will be to greater human uniformity and the result will be a human race that is significantly less biologically resilient and robust.

The potential for germline transformation to result in greater vulnerability to disease is not purely hypothetical. The full phenotypic implications of our genes are demonstrably more complicated than a simple worse to better continuum. Some genes provide a benefit in one environment and a cost in another. Consider the following two illustrations: sickle cell anemia and melanin. Sickle cell anemia is a recessive genetic disorder; one must receive the gene from both parents to suffer it. At the same time, the gene, even if received from only one parent, grants immunity against malaria. Eliminating the gene, while wiping out sickle cell anemia, would increase vulnerability to malaria. The second illustration is the set of genes for the production of melanin in the skin. If for health reasons (greater absorption of vitamins through the skin), or aesthetic reasons human beings universally chose to

genetically reduce melanin and therefore pigmentation in their offspring, it would leave them more vulnerable to skin cancer.

But these are known dangers. A more serious problem results from unknown dangers. Perhaps, for example, the gene that permits human susceptibility to the common cold also provides a partial immunity to some more serious affliction, which we will only discover after eliminating the valuable gene.

What answers are there to this threat to the gene pool from germline engineering? First, and probably most important, before germline engineering could so reduce genetic variety as to present a danger to the continuation of human civilization or even to the entire human gene pool, it would have to be nearly universal. That is a prospect so far in the future, if realizable at all, that it should raise no bar to any current steps down that trail. We are like someone planning a vacation to the Grand Canyon. A concerned friend tells us that if we walk close to the edge of the canyon we face the danger of plummeting to our death. Exercising proper caution does not require that we cancel our vacation or even slow our drive, but rather that we enjoy the view and keep a safe distance from the edge when we arrive. Similarly, if in a thousand years the human race, having eliminated Tay-Sachs, Huntington's and a variety of other genetic diseases, and having raised average human intelligence, is confronted by the prospect of seriously challenging human genetic diversity it should face that problem with the tools then available, not stand in the path of genetic progress now.

Second, we should already recognize and appreciate the power of one tool that will be available to us. Germline engineering is nothing more than the technology of replacing a sequenced string of four chemical bases. The process of germline transformation has no single direction to it. We can go back from whence we came and undo our mistakes in the germline – much as nature does.

Third, we should recognize that it will be human beings acting out of their own self-interest who will make use of germline engineering. A threat to human diversity is a threat to each individual human being. Why assume that the individual response to the threat to genetic diversity will be social wealth reducing? That is, why assume that it is in the interest of the race that diversity be increased, and yet in my interest that my offspring be like everyone else? The opposite is more likely the case; I have at least as much to gain from my offspring being different as does the race as a whole.

To illustrate, imagine that there is a choice of genes that leaves one susceptible to either sickle cell anemia or malaria, as more people choose immunity to malaria, more leave themselves vulnerable to sickle cell. The choices of others change the payoff matrix for me. As a potential disease threatens to wipe out a large portion of the population the payoff to my offspring if they

survive increases – they will inherit a larger portion of the world. Thus there is an increase in the expected wealth of those in the minority.[6]

With respect to infectious diseases there is a further reason that as more people create genetic immunities for their offspring less incentive exists for others to follow. The fewer people who are susceptible to a disease the less probability that I who am susceptible will fall victim. This is much like the effect of, and response to, vaccination. When a large proportion of the population are vaccinated the incentive for me to follow diminishes; there are now many fewer people from whom I can catch the disease.

The general point is that there is no tragedy of the commons at play here. As individuals we have the power to make considered choices, and if preserving diversity is a good bet for the species it is an even better bet for each individual.

An aside on cloning

Cloning has been much in the news of late both as a scientific matter and a bio-ethical one, because it is a distant cousin of germline engineering I shall offer a few words on its benefits and harms. Cloning is the process of artificially and purposefully replicating an organism. Therefore, while the prospect of cloning is driven by many of the same scientific discoveries and innovations as germline engineering, it presents a different and arguably narrower set of potential dangers. From an evolutionary perspective cloning is radically conservative. If successful, it creates no new genotype. Thus, unlike germline engineering, it cannot provide a systematic advantage to a member of the next generation or pollute the gene pool. The harm that cloning might cause comes in two forms. First, given its less than stellar record of success thus far when attempted on other species there is potential harm from failed attempts to clone humans. Second, there is a postulated harm to the clone of being a clone.

In the popular imagination cloning is some sort of Brave New World science fiction enterprise in which there are multiple copies made of some outstanding – or infamous – human being. Would that that were a real prospect! While the notion of cloning a hundred Hitlers made for an entertaining movie plot, it is not likely to hold an attraction even to hard-core Nazis. It is not that Nazis would not like to revive the Führer, it is rather that only a total fool would believe that a clone of Hitler would inherit his political philosophy. As for cloning great geniuses that is a different matter. A newly born Mozart would inherit the same musical genius that allowed the first to compose a symphony at seven; a second Gauss would doubtless be a mathematical giant. And, neither would be limited to the accomplishments of their earlier representation; each could stand not only on their own shoulders, but on those of the geniuses of the past two centuries. And the new Mozart,

Gauss, Einstein, Kant, Hume, or Smith need not enter the career of his predecessor, but could find expression for his genetic genius in any number of different fields. While some express horror at such a prospect I see no dark side to another hundred Mozarts, another five hundred Gausses, another thousand Einsteins, perhaps even a few more Cohens, both creating and procreating.

As attractive as I find the prospect, the likelihood of this coming to be in a liberal democracy is practically nil. Each clone when born will be an autonomous human being entitled to all the rights of any other person. He will not be the property of anyone, certainly not the scientist who carried out the procedure. So where is the self-interest that would drive such an enterprise? Who would be willing to gratuitously gestate and rear such a clone? Who would be willing to compensate others for doing so? It is hard to see where the private market for these clones would be.

The more likely demand for cloning has three sources that arise from the direct interests of those with a strong personal stake in the genetic material to be cloned. First, there are infertile couples who wish to create a genetic legacy. Particularly when both husband and wife are infertile, the only way they can have a child genetically related to either of them is by cloning. Second, there are couples who wish to replicate a deceased child or other loved one. Third, there are people in medical need of genetically matched tissue. Their interest is exclusively, or principally, in using a clone as a means to harvest such tissue.

None of these motives is frivolous. It requires only a modest measure of imagination and sympathy to appreciate the longing. But that does not necessarily make cloning a sensible or a permissible choice. It is not sensible if the expected cost to the actor outweighs the expected benefit. And, perhaps it should not be permitted if the expected costs to all affected parties outweigh the benefits.

What are the potential dangers and harms of cloning? One set of harms turn on the unreliability of the process. Virtually all innovative medical procedures are risky and dangerous in the beginning. Based on the experience of attempted cloning of other species one can predict that there will be much failure before there is success. To the extent that the cost of failure only burdens the cloned person or their sponsor, no public intervention is called for. In each case those directly affected will be in the best position to weigh the costs and benefits.

But what of the clone? Will they be harmed by failure? While complete failure to initiate biological life harms no-one, the creation of a flawed being is another matter. The most serious hereditary defects usually result in spontaneous abortion, so let us consider that as emblematic of the harm of failure to the clone. What weight should be given to the interests of the aborted

fetus? I can offer no definitive answer. Who knows the interest of the hypothetical fetus in the gamble of life and premature death? That said, I suggest that it would be dishonest for many readers to attach excessive weight to this loss. Honesty demands consistency. The costs one assigns to the harm to the fetus from a spontaneous abortion must be consistent with the costs to a fetus from intentional abortion and with the costs to a fetus from negligent and reckless pre-natal behavior by pregnant women. A fetus is a fetus is a fetus. If harming or killing it carries weight, that weight does not change with the act that gave rise to the harm.

It has become the *de facto* law of the land in the United States and most Western countries that a mother has the unconditional right to cause the death of her unborn fetus virtually at any time prior to birth. Indeed, she can have the price of this procedure paid by the state. And, of those who oppose abortion, none but the most fanatical would treat a woman who chooses to have an abortion as a murderer. They thereby imply that they do not really consider the fetus a complete person entitled to the full set of rights accorded to others. Given these widely shared sentiments with respect to abortion, the death of the fetus cannot carry much weight for most of us. And if there is a greater cost that flows from bad intention, remember that, unlike in the case of an intentional abortion, the purpose of cloning is to produce a living healthy infant. The death of the fetal-clone is not only unintended but directly opposed to the wishes of the parents.

To this point I have only discussed the cost associated with failure. What of success? Is the clone harmed by having been created? Some commentators have suggested that being a clone would be an awful burden. I am skeptical. We all enter this world with a particular genotype not of our choosing. Each of us, like George Bailey, the lead character in 'It's a Wonderful Life', might at some point wish that we had never been born. There seems no a priori reason to think that a clone would have any more reason for this sentiment than anyone else. It is disingenuous to not only wax poetic of some imagined burden on a person who happens to be a clone, but further to treat this burden as so serious as to warrant a prohibition. There is a gross inconsistency in condemning cloning for some imagined harm to being a clone, when at the same time the rest of humanity is free to copulate and procreate willy-nilly, and as a result carelessly, negligently and recklessly bring offspring into this world with severe medical, social and economic disabilities. Note as well that if there is a particular psychological burden to being a clone, then identical twins experience much the same burden, the only difference being that clones are not coincident in time.

As for creating a clone for the purpose of harvesting the tissue, that is a different matter. Here the purpose is not normal procreation, but rather the use of the clone as an instrument. Still, let us not be too quick to condemn.

There is more than a little hypocrisy in condemning as wantonly immoral partially gestating a fetus and aborting it for the purpose of recovering transplant tissue to save a life or restore health and yet permit, indeed compensate and encourage, the same abortion to save the mother the burden of pregnancy. Whether the intent is to abort at some pre-natal stage or allow the clone to become a person and then used as a source of donated tissue, for example, bone marrow or a kidney, the taking of which does not severely compromise the health of the clone, the procedure is problematic and should perhaps not be condoned because it entails treating the clone as an instrumentality rather than (or in addition to) a person. But if such behavior is unacceptable it is not because of any peculiarity of cloning. A couple could seek to produce a non-cloned fetus/child for exactly the same reasons. The propriety of such a strategy whether with a clone or not turns on the questions of: (i) at what stage of fetal development does one constitute a person entitled to the respect and care of one's parents; and (ii) the level of sacrifice parents may impose on one child for the benefit of another.

Conclusion

The concern about the human gene pool in particular, and the wider set of fears about germline engineering in general are puzzling. The central arguments to support them are extraordinarily weak. So why is there such strong objection and anxiety?

I believe this is merely a symptom and manifestation of a deeper unease. There is a pervasive sense of spiritual decay in modern society, and rightly or wrongly this is associated with technological advance. That association is not merely an artifact of temporal association, there is a causal connection. The great advances in technology derive from man's successful attempt to rationalize the world. The fundamental unease, I believe, is rooted in a deep sense that at least one central aspect of the world, our lives, is ultimately not rationally comprehensible. We have a suspicion that the attempt to rationally understand ourselves leads us away from, rather than closer to, a comprehension of what we are about.

Our sense of unease with the modern world translates to a nostalgia for an older, simpler time, when perhaps life was more complete and fulfilling. And so the fears about germline engineering are of a single piece with fears of nuclear energy and computers. But the virtue or vice of germline engineering, like that of nuclear energy and computers before it must rest on its own feet, and not merely be implicitly condemned on quasi-religious grounds as one more evocative representation of the misguided arrogance of modern man having lost touch with his roots.

On whether we are indeed lost, I offer no opinion in this chapter. Whether the great advances of science have resulted from this arrogance I yet again

offer no opinion. But finally, if all these things are true, is it a bad thing that modern science has discovered vaccines against polio and smallpox and entire extended families of antibiotics and analgesics? Here I shall not equivocate. These discoveries are not bad. On the contrary they are immensely good; they relieve suffering, cure disease and permit people to live out a normal life span. So too with germline engineering. It offers the prospect of relieving much human suffering and allowing more of us to fulfill our aspiration of having children who exceed ourselves by our own measures of human worth and value.

Further, as I have tried to demonstrate, germline engineering presents us with only chimerical and fantastic dangers. Prohibition, and even mere regulation, of germline engineering, beyond that already extant for medical experimentation must of necessity result in delay and greater expense. This delay and expense in turn must result in those who would otherwise benefit from a miracle of genetic intervention instead having to do without.

One's genes are as much one's property as any other part of one's body. If there is a reason for the state to interfere with the free use of one's genes it rests on the same basis that the state restricts the use of any other private property; does that use do substantial uncompensated harm to others? I have tried to show that the potential of germline engineering to do such harm is extraordinarily small. Beyond that my hope was not merely to demonstrate that these dangers are imaginary, but in doing so to suggest that the entire category of claims to some new category of property requiring a new set of legal, moral or social rules is unjustified.

Notes

* I am grateful to the Law and Economics Center at George Mason University School of Law for its generous support.

1. The Salish Indians of the American Great Plains provide a parallel example. The Salish were commonly known as Flatheads because some of the neighboring tribes, believing that pointed heads were more elegant and attractive, bound the heads of their infants to change the shape. Presumably pointed heads could also be fostered by some sort of germline intervention.
2. By way of analogy note that slaves in the American south in the early nineteenth century had better diets and higher life expectancies than northern factory workers. The slaveowner, because he owned the capital asset of the slave, had a strong interest in maximizing its value.
3. Consider in contrast the human breeding of cattle and chickens. Its purpose is to produce abundant, tender, tasty and inexpensive meat. Were that the motive for people to genetically transform their offspring, then just as there is a sharp divergence between the interests of the chickens and that of the chicken farmer as to the genetic endowment of the chickens, so too there would be one between the child and its parents.
4. It was until recently believed that human beings carried perhaps 100 000 genes. The current estimate based on the results of the Human Genome Project suggests the number to be perhaps one-third of that (Venter et al. 2001, p. 1304). The characterization of this as an estimate understates the uncertainty as to the number. The science of identifying a gene is still quite primitive.

5. Rose Friedman's brother, Aaron Director, is one of the founding fathers of law and economics.
6. In this particular case there is further advantage to being in the minority. The gene in question provides its benefits against malaria even in the heterozygous state but only leaves one subject to sickle cell disease if both pairs of alleles are present. Thus as the percentage of carriers falls the cost to me of carrying the gene (likelihood of passing on sickle cell disease to my children) falls while the gain (my immunity to malaria) stays the same.

References

Dworkin, R.B. (2001), 'Hard cases for autonomy, respect and professionalism in medical genetics', *Annual Review of Law and Ethics*, **9**, 153–64.

Heller, J.C. (2001), 'Testing the limits of parental choice: using genetic applications to produce disabled children', *Annual Review of Law and Ethics*, **9**, 165–72.

Herrnstein, Richard J. and Charles Murray (1994), *The Bell Curve: Intelligence and Class Structure in American Life*, New York: Free Press.

Venter, C. et al. (2001), 'The sequence of the human genome', *Science* , **291**, 1304–8.

13 The contractual nature of the environment

Terry L. Anderson and Bobby McCormick

Introduction

While it took the most-cited article in the history of economic science, Coase (1960), to open our blind eyes to the problem of social cost, economic analysis of the environment has not progressed much beyond Pigou's (1912, 1920) notion that pollution problems result from a divergence between social and private costs. Viewed through the Pigouvian lens, environmental economics has focused on the static notion of efficiency, with policy prescriptions centered around regulations that dictate efficient outputs or taxes that correct prices for uncompensated costs.

At the heart of this approach is the alarmingly simple but deceptively complex term, 'externality'. In the theoretical world of externality, parties to market transactions fail to take into account the effects of their actions on third parties who bear costs (negative externalities) for which they are not compensated or reap benefits (positive externalities) for which they do not pay.[1] Accordingly, market transactions lead to inefficient outcomes with too much of a bad or too little of a good produced.[2] Again the policy prescription is to regulate quantity or tax transactions, or both.[3]

The externality focus in environmental economics implicitly assumes a structure of property rights without ever explicitly recognizing this truism. In the case of negative externalities, the implicit assumption is that the party or parties who bear costs for which they are not compensated have a right to be free from those costs, and in the case of positive externalities that the party or parties who provide the free ride have a right to be compensated for their production. Indeed the mere act of defining an externality amounts to asserting a property rights claim.[4]

In sharp contrast to the paradigm of Pigouvian externalities, Coase taught us that the problem of social costs was best characterized in terms of competing uses for resources for which property rights are not clear (see Yandle 1998). If one person wants to use air as a disposal medium and another wants to breathe that same air, there are competing uses. Those competing uses are not usually resolved through market contracts because property rights are not well defined, enforced and easily transferred, but this is just the result of the cost of defining and enforcing property rights. Conflicts occur, and competing parties have incentives to compete for the rents to the ill-defined resource either by acquiring and capturing property rights to it, by racing to seize them

on the commons thereby creating the tragedy, or by fighting in negative-sum war games.[5] If one person preserves habitat for a wild species, say spawning habitat for salmon, and another can catch the salmon, there is competing use for the resources necessary to produce the habitat, and the party bearing the opportunity cost of those resources, in this simple static story, is not compensated (though he or she might be). Again, market transactions cannot be counted on to regularly resolve the problem because the rights are not well defined, enforced and easily transferred.[6] There may be room for expropriation and self-enforcement of rights, but these would stand outside the normal bounds of the law.[7] In this case, competing parties have incentives to compete for the rents by seeking use rights which cede them the rents. In almost all of these cases, the problem centers around the fact that the property rights are not clear. To assert a negative externality in the first case is to assert a property right on behalf of the person wanting to breathe clean air; to assert a positive externality in the second case is to assert a property right on behalf of the person producing the habitat. To say that air is polluted is to say that a person has a right to breathe clean air. Usually these types of statements are lodged in historical uses or ethical judgments about right and wrong. Somehow it seems natural to breathe clean air, and, as such, there is a natural right to it.

Economic science is not well equipped to resolve who should have what rights, but, because of Coase's insights, it is equipped to consider how different institutional regimes would resolve competing uses. To see how, consider an example where the 'natural rights' are not so clear. Suppose there is a flowing river into which one firm deposits mercury, using the river rather than microfilters, chemicals and trucks to clear the plant of the heavy metal. Suppose another firm downstream has to remove the mercury in order to use the stream water for paper production. In classical Pigouvian analysis, the first firm is said to impose a social cost on the second firm. This line of thinking is wrongheaded. In fact, under common law the rights to the stream are clearly defined. Downstream riparian landowners have rights to beneficial use of the flowing stream that cannot be unreasonably degraded by upstream users. Upstream and downstream users can contract with one another to change stream quality if they desire. Natural rights aside, the problem of competing uses is little more than a recognition of the basic economic problem of unlimited wants and limited means, that is, scarcity. Land, air and water are scarce, and any use of these resources has an opportunity cost. Of course, the resources necessary to define, enforce and transfer property rights are also scarce (see Demsetz 2003). Externality theory assumes away the problem of rights definition by asserting a particular distribution of property rights that do not even exist except by previous use and assumes that regulatory solutions are necessary. Regulation is usually deemed appropriate by

observing or claiming that the costs of enforcing and transferring property rights are prohibitive. In effect, claims about the existence of externalities in this context amount to a delineation of the rents accruing from nature's bounty. Take the case of emissions into the air. To say that air is polluted is to say equivalently, people have a right to breathe clean air. Saying that an externality exists assigns the asset's rents to one particular user or group of users.

The purpose of this chapter is to follow the lead of Coase and focus the analysis of environmental economics on the critical issue of property rights. Recognizing that environmental issues (not problems) are the result of competing uses for scarce resources, we ask how people contract with one another to resolve this competition.[8] In particular, we harken back to Coase's earlier contribution to economic thought, 'The nature of the firm' (1937). There, Coase planted the necessary seeds for the new environmental economics based on property rights. We begin by reconsidering the nature of the firm in the context of contracting costs and continue by considering the types of contracting costs associated with establishing property rights to environmental assets that enable owners of those assets to nurture and capture the rents rather than dissipate them in the tragedy of the commons or in negative-sum battles.

The nature of the firm

Over the past half-century, the classical theory of the firm as it is studied in economics has been transformed by the work of many eminent scholars.[9] At the top of that list is Nobel laureate Ronald Coase who first described the firm as an organization for reducing costs by substituting decisions within the firm for market contracts. Since his seminal article, others have elaborated on his theme and emphasized that firms themselves face agency costs which they try to reduce through internal contracting. The central theme of this major renovation in economic theory is to treat the firm not as a static institution, but as a set of contracts.

The modern corporation probably dates back to the creation of the English East India Company.[10] Early firms were legal creations of the Crown, but over time this role devolved to state bureaucrats. It is important to recognize that in this framework, the firm is a legal fiction. Producing nothing, owning nothing, the firm is merely a theoretical construct heretofore used to economically represent the complex contracts that actually comprise this thing we have come to call the firm. Recognizing this reality focuses attention on individual incentives and the nature of contracts that comprise the modern firm. In this modern version, firms do not do things, individuals do, and they do it through contracting. The firm, as a living team of people, creates internal tensions and opportunities. Managers, representing or working for

owners, can fail in their duty and enrich themselves. Hence independent audits and boards of directors have emerged. Individuals within a firm can specialize in production, recognize efficiency gains in their niche, and save the firm money, but all of this requires either the right incentives or saintliness (the latter being rare). The modern property rights theory of the firm offers a richer, more detailed approach to understanding property rights and responsibility for use of resources. This new approach enhances our understanding of the rationale and value of firms for organizing production and leads to keener insights into managerial compensation, capital structure, dividend policy and complex labor contracts.

The contractual nature of the environment

The potential extensions of this approach have not been entirely exploited, especially when it comes to new applications of the contractual approach to environmental concerns. Whether viewed as an input or an output, the environment is a general term for natural resources whose value can be nurtured and captured to enhance the value of the firm, especially if the ownership of those resources can be accounted for in contractual relationships. When successful, these contracts convert environmental features into assets, thereby improving economic efficiency.

The property rights literature has shown that rights to assets emerge as they become overused and abused. The enclosure movement in England is heralded in this regard.[11] On the western frontier of the United States, Europeans sought ways to contract with American Indians when the property rights were clear, fought with them when they were not (Anderson and McChesney 1994), and hammered out their own property rights once the rights of Indians were effectively extinguished (Anderson and Hill 1975). Numerous other examples such as fishing rights, gold claims, and tradable emission permits amount to rights to use resources in special ways.

Environmental assets offer a new frontier where property rights can evolve and where contractual arrangements within and between firms can improve resource use, profits and social welfare. The physical environment, air, water, land and other components of these natural resources, is used by humans to meet a multitude of demands – life itself. In the absence of some property rights to these resources, competing demands typically lead to overuse in the form of what we have come to call air and water pollution and excessive exploitation of forests, fisheries and so on. To eliminate the tragedy of the commons, some system of rights delineation and contracting must evolve to limit use and increase the value of environmental assets.[12]

Moreover, once use is limited, free-will exchanges via markets are sure to follow. The initial or starting allocation of property rights are rarely efficient once time passes, and therefore any trades stand to improve allocation and

operational efficiency. If rights are inalienable, production and consumption inefficiencies usually follow.[13] Because inalienable rights are typically the devil's cousin, trading is the natural, efficient and inevitable antidote, save for the iron hand of the law.

Our position here is that organizations, which can develop well-defined, enforceable and alienable rights to the environment, will be richer and more valuable. The environmental entrepreneur can increase profits by acting as a proactive creator of value by recognizing that there is an undefined or abused asset and by defining rights to unowned, unused and underpriced assets, the entrepreneur can capture their increased value. The key to entrepreneurship is to find ways to privatize environmental assets and create value where others have yet to recognize the potential. In a word, contracts are the engine that make the value-creating conversion take place.

Identifying environmental assets and establishing ownership claims to them raises several questions. How is the entrepreneur with vision compensated? How is the value of the new product or service measured? What are the opportunity costs of production? How are the rents from value creation to be distributed? And, who bears costs when rights structures shift? It is normal for teams within firms to produce output. If a team recognizes a new and valuable environmental opportunity, how are the members to share in the outcome? Given the problem of dividing up the resource, will rights naturally evolve from the bottom up or will they have to be created by central authority?[14] At issue is what are the contracting costs between individuals and through collective action?

Contracting costs for environmental assets

Four basic contractual costs must be overcome to create environmental assets out of environmental tragedy. These are the cost of

1. discovering the asset's value,
2. defining or measuring the physical nature of the asset,
3. monitoring the use of the asset so that its value is not diminished in the way it is presently used, and
4. knowing the value of the asset in alternative uses so that it can be reallocated if there are higher-valued uses.

Firms must account for all four types of contracting costs and create structures that recognize and reward environmental entrepreneurship and encourage measuring and monitoring the use of environmental assets.

To provide a context for understanding these four contracting costs, consider the case of International Paper (IP) converting its lands primarily used for timber production into wildlife habitat markets through recreational op-

portunities (see Anderson and Leal 2001, pp. 66–70). Value is created in this conversion, but only to the extent that underlying contracts, explicit or implicit, are specified. In the IP case, this conversion first required entrepreneurial vision on the part of biologist Tom Bourland who recognized the potential to increase profits from the asset. In effect, someone in the firm realized that there was a market for the standing trees as a hunting resource.[15]

Once recognized, the entrepreneurial vision for IP lands had to be converted into profits. This required measuring the opportunity cost of forgone wood fiber production and implicitly contracting within the firm for timber and land inputs that were being used for timber production. Were the winners to compensate the losers? If so, what would the currency be? A simple taking of the timber asset away from lumber and paper production would surely create organizational tension. The logging division would complain that its wealth had been confiscated. Some contract or compensation was required to maintain efficient organizational harmony and operation. From the perspective of an input manager, a taking is a taking, whether by government or by higher-level management within the firm. Importantly, unless the logging manager in some sense views the timber holdings as his/hers, he/she will not be the proper shepherd. The manager will cut the trees too soon, and will not offer enough fire protection, or fertilize or innovate in tree silviculture, or have the proper incentive to develop new stands of trees.

There has to be a sense of ownership within the corporation to resolve these agency issues. In its absence, conversion of timberland to hunting land can destabilize the entire corporate structure. The manager of the firm who does not anticipate this discord is likely to suffer. Indeed establishing rights to timber may actually spur the logging division to find alternative higher-valued uses for the timberlands. While the owner of the firm can simply dictate a reallocation of assets within the firm, some form of trading may make the transition smoother, and the form of trading will vary from case to case.

With an internal structure for converting land from traditional wood fiber production to amenity production, the entrepreneur had to develop contracts with new users, monitor their use, and be on the lookout for even more profitable uses. For example, if energy was discovered under a duck hunting pond, the opportunity cost of disrupting the hunting asset would have to be considered, and this would require new contracting within the firm between recreation managers and energy producers. In short, the transformation of IP assets from traditional to new uses required a focus on all the contracting costs listed above. Let us now consider each in more detail.

Discovering the new values

The entrepreneur within this structure is the person who upsets the apple cart. He or she is the one who sees value that others have not seen and tries to reallocate resources already under the control of the firm and/or to combine existing resources with new resources so as to increase profits. Exploiting this newly recognized value requires obtaining control of resources. To the extent that the control is vested with other decision makers, the entrepreneur will have to convince those decision makers that reallocation makes economic sense. This may be done in a hierarchical way, whereby superiors direct others to reallocate or through contractual arrangements within the firm. The former is more likely to create conflict with existing managers who consider themselves owners where the latter requires positive-sum games between the entrepreneur and existing resource managers. C.E.O. John Browne's trading scheme within British Petroleum (BP) to reduce carbon emissions is a perfect example of how the goal might be dictated from above while the means are achieved internally through trading. Again, as noted earlier, discord can and probably will surface from the potential takings or loss of use rights. Only a skillful hierarchical structure will pay real dividends here, but the costs cannot be ignored.

Another example where an environmental entrepreneur recognized the potential to avoid conflict through compensation comes from the case of the Defenders of Wildlife wolf compensation fund (see Anderson and Leal 2001, pp. 172–3). Environmental entrepreneur Hank Fischer recognized that livestock owners would be bearing the cost of having free-ranging wolves. By developing a privately funded program to compensate livestock owners for their losses, Fischer essentially was contracting for habitat. Though Defenders does not own the wolves and therefore is not technically liable for them, they have acknowledged the right of livestock owners to be free from predation.

There must be some structure within the firm to switch on and reward entrepreneurship. Economic models of the firms generally conclude that the entrepreneur is the residual claimant or owner because rewarding entrepreneurship is so difficult. Typically, therefore, it is the chief executive officer (CEO) who takes entrepreneurial risks and bears the costs or reaps the benefits therefrom. However, as firms increase in size or complexity of operation, it is difficult for only a few senior managers to see all the entrepreneurial opportunities and act on them. All sorts of information regularly emerges within the bowels of organizations, and a rights or contract structure that offers incentives and rewards for development of these insights is necessary to create environmental value within the firm. Firms must establish opportunities for others within the hierarchy to take some entrepreneurial risks. Rarely is the valuable information solely in the hands of those at the top. The

key is to find ways of rewarding the good and penalizing the bad. Ultimately this requires a structure that is malleable and flexible enough to allow for the creation and assignment of rights to new environmental assets.

Rewards (and penalties) must be linked, within the order of the firm, to entrepreneurial visions. If individuals have responsibilities, but not rewards, chores go unfinished. If rewards are not pegged to responsibilities, core discipline fails, ultimately, for lack of justice and economic fairness. Consider the classical organizational hierarchy of a complex vertically integrated, multiproduct firm. The CEO or management team sets the rules for operation, the economic and quasi-legal structure of the firm. The day-to-day decisions and contracts are written by the parties directly involved, with oversight and ultimate control by the team at the top. Herein lies the rub; new information will only arrive at the top by accident. Insights, revelations and creative ideas can pop up anywhere within the organization. If this information has to be communicated, usually in full and surrounded by context, all the way to the top of the hierarchy, agency costs may cause a loss of information in the transmission.

The perfect management team is the one that creates an efficient operating environment where every actor within the structure feels all costs and benefits to the firm by every action and decision made on a daily basis. This is impossible, but it emphasizes that the problem is one of designing a dynamic, fluid structure capable of discovering new assets and rights. Because information can accrue at any point in the hierarchy, a centralized decision structure stands to lose some valuable ideas. On the other hand, a diffuse structure that does not properly reward and motivate will suffer agency costs. Success comes from creating a rich, stable structure that can withstand new information about new assets within the context of existing rights and contracts.

When and how quickly environmental entrepreneurship gets switched on will partly be a function of the value of the environmental asset in question. If environmental assets are not very valuable because they are abundant, it will be harder for the entrepreneur to convince others that it is worth taking action to capture the value. For example, if fish are so abundant that fishers can catch all they want without any apparent decline in the productivity of the fishery (that is, there is no apparent tragedy of the commons), it is unlikely that fishers will do anything to privatize the assets or otherwise reduce fishing pressure. However, as more and more fish are harvested, their value and the cost of catching them will increase, providing more return to establishing rights to the fishery. By this time, however, it will be harder to privatize the resource because more people will have a vested interest in the rents.[16] War is not an unusual outcome in situations like this.

Returning to the IP case, we can understand this problem. As long as there are abundant recreational opportunities available without crowding,

hunters, fishers, hikers and campers will not be willing to pay much for the experience, and IP will not be willing to expend much to capture the value of the environmental asset. However, as the value of recreation and other complementary outputs such as endangered species habitat increases, it will be increasingly valuable for the company to privatize and capture this new value. This helps explain why IP has been successful with this program in the Southeast, where there is a tradition of fee hunting and not much free hunting on public lands, while timber companies in the Pacific Northwest have been less successful in the presence of zero-priced public recreational opportunities.

The path and timing of entrepreneurial action with respect to environmental assets will follow the well-known path of innovations, wherein adoption of innovations or new techniques typically proceeds slowly at first by dare takers and creative managers (for instance, see Griliches 1957), gathers momentum as the virtues of the new idea are incorporated by the risk averse managers, and finally reaches a saturation adoption rate. Think of personal computers as a modern example, or in the context of environmental assets, think of marketable or tradable emissions rights. Consider the slowly emerging private water rights to underground water in the American East. Long saturated with freely flowing surface and underground water, population pressure has created scarcity out of abundance. Riparian rights structures are incapable of handling this problem of scarcity, and markets are starting to emerge. Whether the East will become the West in terms of private water rights is yet to be seen, but the process has begun. And, as the East adopts private rights to water, will it imitate the mistakes of Western water law, notably inalienability?

Measuring inputs and outputs

Once entrepreneurship is switched on so that the entrepreneur searches for, finds and attempts to exploit new opportunities, the entrepreneur must measure or specify the inputs that need to be reallocated. In the IP case, Tom Bourland had to secure the necessary timberland to supply the recreational experiences demanded by those willing to pay. On the input side, this meant measuring the size of the parcels, estimating the opportunity cost of the timber not cut, securing the budget to install fences and gates to regulate entry, and so on. On the output side, it meant defining a service that could be marketed, and the main one for IP was hunting. Bourland had to specify how many people would be hunting where, what species would be hunted, what rules would govern ingress and egress, and what price would be charged. Only with all of these attributes of the inputs and outputs defined and measured could the company turn the recreational component of land management into an asset.

Doing this for other types of environmental assets may not be so easy. In the case of carbon emission reduction, BP had to define a unit of reduction (the output), measure the output and monitor each production unit, but it left the determination of how the reductions would be achieved to internal contracting between production units. Applying this to other types of water and air pollutants requires specifying the pollutants that will be reduced, specifying the production process that will be used to achieve this reduction, measuring the opportunity cost of achieving the reduction, and determining whether there is an asset value in the reductions.

Consider the case of high voltage electricity transmission towers and their impact on large birds of prey. Some argue that power lines can shock and kill birds such as eagles and similar large wingspan fowl. At the same time, transmission towers provide nesting opportunities. Under the appropriate rights structure, the transmission towers can become environmental assets. The company owning the tower can offer to market raptor nesting space on its towers to environmental groups willing to pay for nesting space. The company might offer tours of towers where eagles nest. It could offer to build safer towers for predator protection if a demander were willing to pay. Armed with the right incentive structure, the towers can be converted to an environmental asset, but, of course, all of this is subject to contracting costs.

Monitoring and pricing inputs and outputs

Once new values are recognized and measured, decision makers must gain control of the necessary resources to produce the new values. Essentially they must either have the authority to command the new use or contract with others who currently control it. In either case this requires monitoring resource use to ensure that the reallocation actually takes place.

Again, the IP case is illustrative of the costs associated with monitoring and pricing inputs and outputs. Internally, the company had to monitor the actions of timber harvesters so that they did not interfere with production of hunting. Externally, it had to monitor the customers to exclude non-payers and to make sure that they were abiding by the rules. Fences and gates, combined with trespass laws and policing authorities, were necessary to carry out this monitoring. These monitoring costs were reduced by self-policing hunting clubs, which did not want to share their territories with non-payers.

Monitoring costs can be reduced to the extent that there is a link between responsibility and rewards for individuals who have a right to say no to a reallocation, the corollary of fee simple ownership. Decisions must be made by individuals with responsibilities and rewards inside the ordered structure of a firm. For example, if the environmental engineer is responsible for a clean workplace or the safety engineer is responsible for an accident-free workplace, then these people must have charge over decisions about cleanli-

ness and safety. Moreover, their pay and compensation should, at least in part, be tied directly to operational cleanliness, safety, or environmental quality within the firm. In the absence of these incentives, demands to 'clean up' or 'be safe' are hollow chants, by and large, and while they may set a tone for care, they cannot, any more than voluntary income tax payments, be counted on to carry the load.

Pricing environmental inputs and outputs is more costly to the extent that they are more entrepreneurial. If the product is entirely new, it is costly to know what people are willing to pay. What is a day, week, or season of hunting worth? How does that value vary with the number of others hunting in the area? How does it vary with success? What are the recreational alternatives? All of these are questions that must be answered before the entrepreneur can market his/her product.

Take recycling as an example. Some recycling is economically efficient; under the right circumstances it can pay to collect and recycle aluminum, but others, newsprint for example, make little direct economic sense. The costs of collection exceed the revenues. Hence, having a recycling manager may or not be a good idea, depending upon whether he or she has rights, responsibilities and rewards linked and aligned with corporate objectives. An environmental resource manager, as a special position within the hierarchy of the firm, can be useful, but probably more as an educational specialist rather than a creator of rules. Each manager, properly motivated and rewarded, should be the environmental resource manager. The environment is not special in this regard.

Knowing the opportunity costs

A final transaction cost that must be considered is the opportunity cost of devoting inputs to a particular production. Here, it is particularly important for entrepreneurial remuneration to be tied to the residual claim. In general, economists have concluded that entrepreneurs must be residual claimants because of the high cost of measuring and monitoring their contribution to team production (see Eggertsson 1990; Barzel 2003). As residual claimants, entrepreneurs have an incentive to continually monitor the opportunity costs of inputs and ask whether those inputs would generate a higher return if reallocated. Whether this be more of the firm's budget devoted to improving safety in the workplace or more land devoted to recreation, the entrepreneur must be aware of the opportunity cost of the inputs. If the entrepreneur does not have a claim on the gains to be had from reallocation, he/she has less incentive to accurately measure the opportunity costs of the alternative uses or to reallocate to those higher valued uses. Conversely, if the entrepreneur does have a claim on the residual from improved reallocation, he/she has every incentive to continuously consider opportunity costs.

For instance, IP resource managers had to decide whether to continue devoting land to waterfowl hunting or to explore for and develop expected energy reserves under the waterfowl habitat. If the entrepreneur who discovered the value of waterfowl hunting has a claim on the residual value of reallocating, he/she has an incentive to carefully consider the opportunity cost. Otherwise, the entrepreneur is likely to resist because there is nothing to be gained personally from the reallocation, and in fact, his/her contribution to the team production may actually decline if the resources are taken away and reallocated to another.

The importance of residual claimancy to switching on entrepreneurial awareness of opportunity costs is illustrated by the contrast between the Audubon Society's willingness to explore for and develop energy in wildlife habitat that it owns, but not on public lands. In the debate over the Arctic National Wildlife Refuge (ANWR), the Audubon Society and other environmental groups have been a vociferous opponent. Though many people argue that there are innovative, environmentally sensitive ways for exploring and developing ANWR, environmental groups including Audubon have no reason to consider the opportunity costs of just saying no. In sharp contrast is Audubon's history of allowing energy development on its privately owned preserves. In the Rainey Wildlife Sanctuary, a privately owned Audubon preserve off the Louisiana coast, the Audubon Society allowed energy production until the reserves were depleted. Because it was concerned about the wildlife habitat that it owned and because it could say no to drilling if not done in environmentally sensitive ways, the society required extra precautions, but it still allowed energy production. To have said no completely as it has in ANWR would have carried a high price tag, costing Audubon over $1 million per year in forgone royalties. Being a residual claimant, Audubon had a claim on the increased value from allowing energy development, but it also had an incentive to consider the cost to wildlife habitat. Ownership got the incentives right and forced the Audubon Society to consider opportunity costs.

An accounting system, free flowing with price and value information, must exist in order for environmental assets to emerge. The great challenge for the senior management team is to design the correct rights and incentive structures and then to install the information monitoring system that, in effect, runs itself. Information and incentives are therefore fundamental to success in creating environmental assets.

Contracting outside the firm

Even if the entrepreneur inside the firm can figure out how to capture additional value from natural resources controlled or owned by the firm, how can he/she capture value outside the firm? Asking this question goes directly to

how important the Coasean lens is to environmental questions. To assert that there is a negative externality, say from the use of the air as a disposal medium, is to recognize that someone is not capturing the rents from owning and marketing air. The company that overcomes internal contracting costs to reduce energy use, and thus reduce disposal of waste into the air, benefits from reduced energy costs. At the same time, however, it produces a free lunch for those who want to use the air for other purposes such as breathing or viewing. If that company could require others who want to use the air for breathing or viewing to pay for the higher-valued asset, it would have an incentive to remove more emissions.

Looked at the other way, as Coase would, an owner using the air for breathing or viewing might enforce his or her property rights by charging any emitter for using air as a disposal medium. Such a charge would provide an incentive for the emitter to optimize emissions. The free lunch would remain to the extent that other breathers and viewers can now enjoy a higher-valued asset as a result of the owner's enforcement effort. If the owner could require others who enjoy this asset to pay for its higher value, he or she would have an incentive to spend more on enforcement and emissions reduction.

The entrepreneur who can overcome the transaction costs associated with defining and enforcing property rights to air can then market that resource to people wanting to use it as a disposal medium or to people wanting to breathe it or look through it. In the IP case, it took entrepreneurship to overcome the internal contracting costs to reallocate the use of timberlands owned by the firm. It also took entrepreneurship to overcome the contracting costs associated with marketing the recreational amenities produced with the reallocation.

Considering environmental issues through the Pigouvian lens of externalities not only implicitly asserts property rights where they are unclear, it implicitly assumes that the transaction costs associated with definition, enforcement and exchange are prohibitive. Though we would be the last to argue that transaction costs might not potentially prohibit contracting, we emphasize that transaction costs are a cost like any other, such as production, transportation or information costs (see Demsetz 2003); they drive a wedge between potential gains from trade. It makes no more or less sense to say that gains from trade will be greater if transportation costs are reduced than it does to say the same about transaction costs. The question is: can the political process reduce these costs and improve efficiency?

Regulations that determine how environmental resources will be used implicitly assign property rights to one user or another, and in so doing, they create a new set of contracting costs. Like internal contracting costs, these external transaction costs will affect resource allocation. In other words, going from the argument that contracting costs prohibit property rights formation and Coasean bargaining to the argument that quantity or price

regulations will improve resource allocation potentially ignores the different set of transaction costs that arise with regulations.

Of importance to this discussion, local, state, or national governments establish rules and regulations that can stand as barriers to environmental asset value creation. First, environmental regulation often precludes alienable rights structures. The notorious case of air pollution regulations illustrates this problem. Scrubbers are mandated by the US Environmental Protection Agency (EPA) to limit air pollution output of coal-fired electricity power plants. It is well documented that these scrubbers are technically and economically inefficient (see Ackerman and Hassler 1981). However, if AEP, one of the nation's largest coal-fired electricity generators, wanted to uninstall the expensive scrubbers, import low sulfur western coal, emit less SO_2 and NO_x, and collect a reward from the New York City Free Market Clear Air Coalition, it would be precluded by an act of US Congress and EPA regulations.[17]

Because the external iron hand of government can limit and restrict the capacity of the firm to create environmental assets and nurture the rents therefrom, environmental entrepreneurs focus some effort on manipulating the political process. In order to capture environmental asset value, entrepreneurs must be able to improve, protect, nourish, guard and trade their assets. This often requires lobbying to change the political landscape to allow ownership and trading.

Conclusion

If environmental economics is to be extracted from its externality rut, it will have to be grounded on the same foundation of modern understanding as the firm, namely property rights and contracting costs. Accordingly, we should expunge the concept of externality from the environmental literature and replace it with property rights and contracts. The former assumes away the problem by asserting property rights that may or may not exist and generally assumes a solution by asserting that contracting costs prohibit enforcement and exchange of property rights.

By building environmental economics on the same foundation that has allowed us to better understand why we have firms and how they operate to allocate scarce resources, we can better understand how to produce environmental assets. This requires thinking about environmental assets as outputs that require coordination of inputs if they are to be produced. Because this is what markets and firms are all about, this approach will allow us to better understand whether and how this production will occur in the marketplace, and if not in the marketplace, how production can be facilitated in the political arena.

If entrepreneurship is to be switched on in the marketplace, individuals making decisions must feel the consequences of their decisions. This state-

ment, accepted by most managers with respect to finance, labor, safety and so on, seems to fall through the cracks when it comes to environmental assets. There are several reasons for this. First, environmental assets have not always been so valuable. As long as resources are abundant, it is not worth defining and enforcing property rights to them because definition and enforcement efforts are costly. Only when entrepreneurs perceive the asset value to be high enough are they willing to undertake these costs. Second, many people have conceptually viewed the environment as a public good, not assignable to any person or entity. Environmental assets may not be owned and conserved because it is technologically impossible to define and enforce rights to the environmental asset in question or because it is institutionally impossible to make that assignment. The latter especially happens when laws will not allow entrepreneurs to privatize environmental assets. Finally, firms may not take account of environmental assets because the incentives that switch on entrepreneurship within the firm may not be in place. Contracting costs within the firm are not zero, making it more difficult to contract for the production and sale of environmental outputs.

Step one in the evolution of good environmental asset management within the firm is to reduce contracting costs and get the incentives right. This requirement seems obvious in the context of the corporate jet, the copier machine, the company logo, or safety in the workplace, but more abstract and confusing in the context of company air or water quality or land resources. By focusing on the transaction costs associated with measuring and monitoring the use of environmental assets in the same way that we focus on these costs for other aspects of firm management, we can begin to explore the possibility of making the environment an asset rather than a liability for the firm.

Notes

1. Bator (1958) among others has gone to great lengths to try to distinguish between what he calls pecuniary and technological externalities. The first of these being where one party impacts market prices to the detriment of another and the second where one party consumes a resource to the detriment of another. Presumably the first does not distort the market and create an inefficiency or social problem while the second does. Of course, to the 'harmed' party, both create costs that he or she would prefer to not have. Implicitly, the Bator dichotomy assumes that people do not have property rights to prices, but they do have property rights to resources. Our emphasis here is that it is property rights, not costs, that are the key to better understanding environmental issues.
2. Many people worry not so much about the resulting inefficiency, but about the resulting income distribution and economic fairness of a system where one party pays and another one reaps.
3. The concept seems so simple and obvious that, at least early on, few even challenged the idea. See, for instance, the discussions in Meade (1952) and Bator (1958), later corrected by Cheung (1973), where economists are somewhat to blame for converting logic into reality without sufficient due diligence.
4. Though not quite put in these terms, this point was made by Dahlman (1979).

5. For a discussion of these alternatives in the context of Indian–white relations on the American frontier, see Anderson and McChesney (1994).
6. We leave aside here, at least momentarily, the idea of the stream owner striking a deal with the fisherman to create habitat.
7. One can think of a thug beating up those who dare fish on 'his' lake. Here rights might be defined, but not in the classic sense. Nor in this case would we think that these rights are alienable. Albeit more efficient than the commons case, this is probably not as efficient as fee simple rights. The question that remains in all these discussions is: what is the cost of defining and enforcing rights?
8. Saying that there are environmental problems is, in our minds, no different from saying there are transportation problems because chickens won't fly from farms into frying pans.
9. See, for example, Coase (1937), Alchian (1969), Alchian and Demsetz (1972), Klein et al. (1978), Williamson (1975), Jensen and Meckling (1976), Barzel (1982, 1997), to mention a few.
10. See Anderson et al. (1983) for more details on the creation of the corporation.
11. See the classic paper by Hardin (1968) for some more discussion on this point.
12. This issue is complex. Why is it that rights have to be created and enforced by what we normally call government? Why can't bullies simply create and define rights? Are kings governments? Is a family a government? Do private land developments, such as gated housing areas, define rights? Are external monitors and court systems necessary for private land arrangements to function efficiently?
13. A particularly relevant example of inefficiency resulting from inalienability comes from water. Restrictions on transfer of water rights from diverted uses, such as irrigation, to instream uses, such as fish habitat, prevent efficiency gains. See Anderson and Snyder (1997).
14. The nascent Napster example provides a contemporary problem. Water and air emissions offer an environmental example. In these cases technology provides a way of resolving property rights. Like barbed wire, for example, chemical markers in emissions can help define and enforce rights. See Anderson and Leal (2001).
15. Other examples might include owners of historic homes realizing that tour income might exceed implicit rental value of a house. The Biltmore Mansion and the Hearst Castle come to mind.
16. For a complete discussion of the evolution of property rights in this setting, see Anderson and Hill (1975).
17. The same politics pervades the diesel engine emission rules promulgated by the EPA in January 2001. This rule is to reduce emission standards for 2007 and subsequent model year heavy-duty diesel engines (66 FR 5002, 18 January 2001). These emission standards require a 90 per cent reduction of nitrogen oxide emissions, 72 per cent reduction of non-methane hydrocarbon emissions, and 90 per cent reduction of particulate matter emissions compared to the 2004 model year emission standards. The standard has no provision for emission reduction other than a strict control on engine design. No amount of entrepreneurship within the firm can overcome this iron hand.

References

Ackerman, Bruce and William T. Hassler (1981), *Clean Coal/Dirty Air*, New Haven, CT: Yale University Press.

Alchian, A. (1969), 'Information costs, pricing and resource unemployment', *Economic Inquiry*, **7**, 109–28.

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

Anderson, G., R. McCormick and R. Tollison (1983), 'The economic organization of the English East India Company', *Journal of Economic Behavior and Organization*, **4**, 1–7.

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** (1), April, 163–79.

Anderson, Terry L. and Donald R. Leal (2001), *Free Market Environmentalism*, revised edn, New York: Palgrave, St. Martin's Press.
Anderson, T.L. and F.S. McChesney (1994), 'Raid or trade: an economic model of Indian-White relations', *Journal of Law and Economics*, **37**, 39–74.
Anderson, Terry L. and Fred S. McChesney (2003), *Property Rights: Cooperation, Conflict, and Law*, Princeton, NJ: Princeton University Press.
Anderson, Terry L. and Pamela S. Snyder (1997), *Water Markets: Priming the Invisible Pump*, Washington, DC: CATO Institute.
Barzel, Y. (1982), 'Measurement costs and the organization of markets', *Journal of Law and Economics*, **25** (1), 27–48.
Barzel, Yoram (1997), *Economic Analysis of Property Rights*, New York: Cambridge University Press.
Barzel, Yoram (2003), 'Property Rights in the Firm', in Anderson and McChesney (eds), pp. 43–57.
Bator, F. (1958), 'The anatomy of market failure', *Quarterly Journal of Economics*, **72**, 364–79.
Cheung, S.N.S. (1973), 'The fable of the bees: an economic investigation', *Journal of Law and Economics*, **16**, 11–33.
Coase, R. (1937), 'The nature of the firm', *Economica*, **4**, 386–405.
Coase, R. (1960), 'The problem of social cost', *Journal of Law and Economics*, **3**, 1–44.
Dahlman, C.J. (1979), 'The problem of externality', *Journal of Law and Economics*, **22** (1), 141–62.
Demsetz, Harold (2003), 'Ownership and the externality problem', in Anderson and McChesney (eds), pp. 259–81.
Eggertsson, Thráinn (1990), *Economic Behavior and Institutions*, New York: Cambridge University Press.
Griliches, Z. (1957), 'Hybrid corn: an exploration in the economics of technological change', *Econometrica*, **25**, 501–22.
Hardin, G. (1968), 'The tragedy of the commons', *Science*, **162**, 1243–8.
Jensen, Michael and William Meckling (1976), 'Theory of the firm: managerial behaviour, agency costs and ownership structure', *Journal of Financial Economics*, **3**, 305–60.
Klein, B., R. Crawford and A. Alchian (1978), 'Vertical integration, appropriable rents, and the competitive contracting process', *Journal of Law and Economics*, **21**, 297–326.
Meade, J.E. (1952), 'External economies and diseconomies in a competitive situation', *Economic Journal*, **62**, 54–67.
Pigou, Arthur C. (1912), *Wealth and Welfare*, London: Macmillan.
Pigou, Arthur C. (1920), *The Economics of Welfare*, London: Macmillan.
Williamson, Oliver (1975), *Markets and Hierarchies: Analysis and Antitrust Implications*, New York: Free Press.
Yandle, Bruce (1998), 'Coase, Pigou and environmental rights', in Peter J. Hill and Roger E. Meiners (eds), *Who Owns the Environment?*, Lanham, MD: Rowman & Littlefield, pp. 119–52.

14 Government regulation and property rights

Dwight R. Lee

Introduction

I begin my comments on government regulation and property rights by emphasizing one of the least-noticed insights from one of the best-known quotations in economics. Adam Smith ([1776] 1981, p. 56) famously states, '[Every individual] generally, indeed, neither intends to promote the public interest, *nor knows how much he is promoting it.* … He intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which is no part his intention' (emphasis added).

Economists, and others, when discussing Adam Smith's invisible hand, have emphasized how it renders good intentions unnecessary to achieving socially beneficial results. No one denies the importance of this insight, or faults the emphasis it has received. But to fully understand the implications of the invisible hand for a wide range of issues, including 'government regulation and property rights', one must consider the part of the Smith quotation I have italicized – when self-seeking people promote the public interest under the guidance of the 'invisible hand', they do not realize how much they are promoting it.

On the other hand, one could describe the intentions and effects of government regulators by paraphrasing Smith's famous statement as follows: 'Each government regulator generally, neither intends to harm the public interest, nor knows how much he is harming it. … He intends (more often, pretends concern for) only the public interest, and he is in this, as in many other cases, led by an invisible hand to promote an end which is no part his intention'.

With these two quotations as background, I shall make two related arguments. First, the enormous benefits from markets, benefits diminished by regulation, are not easily noticed (the real Adam Smith quotation), and even when they are, few understand their dependence on private property. This first argument goes a long way to support the second one (and the paraphrased quotation); that is, the harm from government regulations is seldom intended or noticed, so there is little resistance to expanding them even though the harm exceeds the benefits.

I shall develop the first argument, emphasizing the necessity of private property for the proper functioning of the market, in the next section. Then, I consider the second argument by presenting a public choice perspective on the importance of political incentives to the enactment, implementation and

consequences of government regulation. Here I discuss the general tendency of government to impose regulations that, by undermining private property rights, harm economic performance and reduce social welfare. Subsequently, I present examples of the harm from government regulations designed and implemented with a callous disregard for property rights. A final section concludes.

Private property and the blessings of the market process

Without private property the market process would be fatally crippled. But the benefits of the market are easily taken for granted, and the connection between those benefits and private property is seldom recognized. This leaves the market process vulnerable to assaults against property rights motivated by ignorance and avarice.

Communication and cooperation

Our prosperity and freedom are made possible by a network of communication that allows widely dispersed people with little direct knowledge of, or concern for, each other to join in a cooperative effort that serves their diverse interests with unparalleled efficiency. Private property rights are at the heart of this network. When these rights are well defined and enforced, property is exchanged at prices reflecting its highest-valued use. Through these market prices people communicate clear, accurate and constantly updated information to each other on the values they place on the resources they own and those owned by others.

Market prices not only convey information. They target the information to those who can make the best use of it, and provide the motivation for them to do so. Millions of people are simultaneously exchanging information through market prices on countless goods and services, ensuring that the mix of goods and services produced and consumed is constantly and appropriately adjusting to changes in relative resource scarcities, weather conditions, political events, technological improvements, consumer preferences and so on.

Furthermore, this amazing degree of market cooperation ensures that each item in the market mix of goods and services is being produced as efficiently as possible. Producing even the simplest product requires coordinating the specialized skills and efforts of many geographically dispersed people using a wide variety of geographically dispersed resources. As Read (1958) explained in a famous article, the materials in a simple lead pencil are so varied, with the many skills necessary to process those materials, bring them together, and assemble them into a pencil requiring so much specialized training, that no one person can make a pencil. Yet, through the coordination from market prices, pencils are produced and made available to us so cheaply, and in just the right numbers, that we never consider what an amazing feat this represents.

And we take more than pencils for granted. How many give much thought to the availability of far more complex items such as televisions, computers and jet aircraft, none of which could be produced as cheaply, with as much reliability, and in the appropriate quantities (if they could be produced at all) without market prices coordinating the decisions of countless people in ways that improve the conditions of all?

It is not only because the market process works so well without conscious direction that we simply accept the tremendous material benefits it delivers with little, if any, thought or understanding. Market benefits are so impersonally and indirectly delivered, and so widely dispersed, that it is almost impossible for individuals to have any direct appreciation for how much their actions benefit others, or how much the actions of others benefit them. And few understand the essential connection between their material prosperity and private property rights which make market cooperation possible.

Private property and freedom

Freedom is necessary for the proper functioning of markets. Without the freedom to enter into the occupation of one's choice, to hire and fire at will, and to buy and sell at mutually agreed-upon prices, the information communicated through market prices would be so distorted that the wealth-creating cooperation of the marketplace would be horribly hampered. This connection between freedom and prosperity has been highlighted by studies ranking countries by prosperity and broad indexes of freedom, establishing a strong positive correlation between the two (see Gwartney and Lawson 2002; O'Driscoll et al. 2002).

But freedom not only enhances the functioning of free markets; private property and voluntary exchange of free markets are essential to maintaining freedom. Genuine rights to private property allow people to exercise a wide range of choices over the use of that property. As Milton Friedman has pointed out, your freedom to express opposition to government, for example, is less likely to be suppressed when paper and printing presses are private property that can be purchased in the marketplace than when they are owned and allocated by the government.[1]

Private property and market exchange also protect freedom by establishing the only setting in which freedom can be tolerated. To be tolerated, freedom has to be exercised responsibly – to take into account the concerns of others. Only in response to the information and motivation communicated through market prices is there any hope that freedom will be widely exercised with such responsibility. I can tolerate the freedom of others to consume as many resources as they choose in pursuit of whatever objectives they may have, only when they do so through markets, because only then will they pay prices for their choices that reflect the marginal costs those choices are imposing on

me, and others. Similarly, others can tolerate my freedom to pursue my objectives in the marketplace because I also will increase my use of resources only as long as they are worth more to me, at the margin, than they are to them.

There is no mystery why people are denied basic freedoms in countries where private property is severely limited and market exchanges are suppressed. Freedom without responsibility is unacceptable, and without markets, people will resort to political coercion to smother freedom.

Even in primarily free-market economies, when private property rights do not exist we substitute political coercion for freedom. For example, excess pollution results from the lack of private property in the atmosphere and waterways, so market prices do not discipline the emissions of waste into the environment. With pollution markets, polluters would have to pay prices that reflect the cost their emissions impose on others, so polluters would voluntarily limit their discharges to something close to efficient levels.[2] But without markets for the right to pollute, we accept – indeed, demand – bureaucratic restrictions on polluting activities that we would consider outrageous in most areas of our lives.

Unfortunately, people are just as oblivious to private property's importance to freedoms as they are to its importance to material prosperity. People seldom appreciate how compromising private property rights (particularly when they are the property rights of someone else, or of a business) undermines the market discipline that allows freedom to be tolerated. Even if people knew that freedom is diminished when property rights are violated, few would realize the extent of their personal loss. As Hayek (1960, p. 32) emphasized, 'The benefits I derive from freedom are thus largely the result of the uses of freedom by others, and mostly of those uses of freedom that I could never avail myself of. It is therefore not necessarily freedom that I exercise myself that is most important for me'. When regulations chip away at freedom by circumscribing the role of private property, few people will notice the decline in freedom, or the loss they suffer as a result.

Dispersed and delayed

People tend not to notice the benefits from the steady improvements in productivity created by the market process because the benefits are widely dispersed and often delayed.[3] For example, assume that a corporation shuts down an unprofitable plant that is a major employer in a small town. How many reports of this event on the evening news will emphasize the net benefits of this shutdown? I have never seen one. Yet such plant closures are part of the adjustment response to price communication that enriches us all by continually directing resources into their highest-valued uses. Indeed, even those who lose their jobs are better off living in an economy in which

such job losses permit appropriate adjustments to changing economic conditions. But few see layoffs, bankruptcies and foreclosures as desirable. Though the benefits from these adjustments are large in aggregate, they are so widely dispersed that the gain for one person from any particular adjustment is too small to notice.

Also, benefits from the status quo – which we have grown accustomed to – are always easier to identify and appreciate than equivalent, or greater, benefits delivered impersonally and indirectly through market processes. For example, how many exporters receiving government subsidies would be willing to give them up for reductions in trade restrictions that would increase their profits by at least enough to offset the value of the subsidy?

Finally, the value from market adjustments may not be realized until much after the adjustments begin. Most workers laid off today will eventually move into new jobs in which they create more value than they did in the jobs they lost.[4] Similarly, the physical resources used by a firm experiencing bankruptcy will be reworked and reallocated to firms in expanding industries where they now create more value. But these movements take time, and so the benefits rooted in private property are both delayed and dispersed, with both effects diminishing, if not eliminating entirely, any awareness of, or appreciation for, private property.

Small benefits trump large costs in the political process

If the political process worked as well as the market process, the opportunity for some to benefit by imposing uncompensated costs on others would be very limited. In this situation, there would be little concern that government regulations would impose costs in excess of benefits, either in total or at the margin. Only if the regulation created benefits greater than costs could those receiving the benefits get the regulation enacted by compensating those suffering the costs. Unfortunately, biases inherent to the political process allow organized groups to capture private benefits from regulation by imposing uncompensated costs on others. The result is a strong tendency for the political process to magnify small benefits while disregarding large costs.

Concentrating benefits and dispersing costs

Government action can, and often does, create benefits concentrated on relatively small groups and paid for by the general public. A government policy that concentrates benefits on a relatively small group animates that group to support the policy politically. The benefits are spread over few enough people to be significant to each, and the group is small enough not to be paralysed by the free-rider problem often associated with collective action. Indeed, the group is probably already organized around a common – typically occupational – interest, so the cost of lobbying for the policy is often quite low.

In contrast, the cost of the policy is spread out over the entire population in the form of higher taxes, higher prices and a less efficient economy. So cost far larger than the benefit is still so small to any one person that it is hardly noticed, if noticed at all. Even if someone does notice the policy's cost, his/her share of that cost would pale in comparison to the expense of actively opposing the policy. So even if someone were assured that his/her opposition would be effective, he/she would still have no motivation to mount that opposition. But effective political action requires a collective effort that, given the overwhelming free-rider temptations facing a group as large as the general public, can be achieved only at an enormous cost, if achieved at all.

Also, even if the cost of a particular program passes some awareness threshold, few, if any, would trace that cost back to the policy that caused it. In part this is because of the difficulty of the task. As explained in the previous section, the connections between the market process and the benefits it provides are not obvious, nor are the connections between government policies that hamper the market process and the full costs of those policies. And even if it were easy to connect the cost of a government policy to that policy, few would do so. Acquiring information is costly and people want only information worth more to them than it costs to acquire. Information on the cost of a government policy is worth almost nothing to individual citizens, given the small fragment of that cost they pay and the even smaller probability that they could affect the cost of the policy even if they devoted all their resources to the effort. As public choice economists have recognized since the work of Anthony Downs (1957, ch. 13), remaining 'rationally ignorant' about most political issues is sensible for most citizens.

Ignorance about a policy is not rational, however, for those who disproportionately benefit from it. First, the connection between the benefits from a government policy and that policy are typically rather obvious to small groups receiving most of those benefits. Indeed, such policies are usually considered in the first place because the beneficiaries, fully aware of the benefits, made sure they were put on the political agenda. Second, regardless of why a policy is being considered, the organized group that stands to realize disproportionate benefits from it will be strongly motivated to become sufficiently informed to make sure it is designed and implemented to increase their benefits.

So a strong bias in the political process amplifies the communication of benefits and dampening the communication of costs. When a politician considers a proposal that benefits the few at the expense of the many, he/she will hear loudly, clearly and incessantly from the few, and little, if anything, from the many. Therefore, government programs and regulations with social benefit/cost ratios less than one (often much less) are enacted and expanded because their political benefit/cost ratios are greater than one.

Enlisting help from the victims

The political bias favoring wasteful policies is accentuated by another distortion in political communication – one that motivates people to support special-interest policies that harm them. This distortion allows politically organized interests to enlist their victims as political allies. Without public support it is extremely difficult, if not impossible, for an interest group to get a policy enacted, no matter how organized it is, how much its members would gain from the policy, or how widely the cost of that policy is dispersed over the general public. But how does a small group enlist public support for a policy that benefits it by harming the public? Unfortunately, it is not difficult and often accomplished. The trick is to obscure the private-interest reality motivating the proposal with public-interest rhetoric. This is often successful for two related reasons.

First, as discussed earlier, most people are 'rationally ignorant' about policy issues since most policies do not cost any one person very much, and even if they did, there is little that person can do about it anyway. So few people are informed on the costs of a policy and, for the same reason, they are not informed on its social benefits (or lack thereof). Thus clever lobbyists and publicists can convince the public that socially harmful policies promote worthy social goals such as protecting American jobs (import restrictions), saving the family farm (agricultural price supports), helping the poor (minimum wages and rent controls), improving public schools (expanding the Department of Education), protecting the environment (expanding the Environmental Protection Agency), and protecting consumers (occupational licensing). People identify with such goals and feel good supporting policies that superficially appear to promote them.

Second, even if people know that a policy would cost them far more than the personal benefit they would derive, a minor feeling of virtue from supporting it at the polls will likely motivate them to do so. Voting allows people to support a policy *expressively*, either by voting for it directly or voting for a politician who supports it, at almost no personal cost. Even if the personal cost a person will pay for a policy if it is enacted is very large, that one person's vote is highly unlikely to make any difference in the election outcome means that the expected personal cost of voting for the policy is very low. Assume, for example, that a policy will cost Tom $1000 if it is enacted. But because in most elections the probability that any one person will cast a decisive vote is miniscule, say one in 20 000 (which is surely lower in most state or national elections), Tom's expected cost of voting in favor of the policy is not $1000, but $1000 × 1/20 000, or a nickel. All it takes to motivate Tom to vote for the policy is for it to give him a feeling of virtue worth at least a nickel.[5]

So by packaging private-interest proposals with public-interest rhetoric, organized interests can motivate public support for government policies harmful

to the public interest – providing small benefits to an organized few by imposing large costs on the unorganized many.

In the two preceding sections I have argued that (i) private property and market exchange receive little, if any, organized political support even though they are the source of enormous wealth and (ii) government regulations receive active and effective political support even though they destroy wealth (certainly at the margin). The discussion in these sections springs from the two quotations in the first section of this chapter, the first by Adam Smith and the second, my paraphrase of the Smith quotation applied to the political process. The public benefits of the market are greatly underappreciated because those who pursue their own interest in the marketplace 'neither [intend] to promote the public interest, nor [know] how much [they are] promoting it. [They] intend only [their] own gain'. So it is easy for people to see the pursuit of self-interest and personal aggrandizement in the marketplace, but remain oblivious to the enormous, and widely diffused, social benefits being created. In the political process, government regulators and those who support them 'neither intend to harm the public interest, nor know how much they are harming it. … They intend (more often, pretend concern for) only the public interest'. So it is easy for people to believe that social concern rather than self-interest is motivating political decisions, and fail to see the enormous, and widely diffused, social harm being inflicted by distorting Smith's invisible hand.

I next consider some particular government regulations that are politically popular despite being wasteful and counterproductive.

Examples of regulations undermining private property

There is unfortunately a large menu from which to choose examples of government regulations that harm society by undermining private property rights. I have chosen three to consider in some detail from the perspective developed in the first three sections. They are mandating employers to provide specified benefits to workers, restrictions on the market for corporate control, and regulation aimed at reducing global warming.

Mandated benefits

Both state and federal governments mandate that employers provide certain benefits to their employees, such as paid leave, health insurance, specified safety equipment and procedures, and retirement benefits. Firms are often exempt from some mandates if they employ fewer than a given number of workers (commonly 50), or meet some other conditions, with the specific mandates, and conditions, varying significantly at the state level. Even when a particular benefit is not mandated, if a firm provides it voluntarily, then mandates commonly apply on how much is provided. For example, there is

no federal requirement that firms provide health insurance to their employees, but if a firm does, many states mandate the extent of the coverage by specifying limits on deductibility and the type of illnesses covered.

Such mandates restrict private property and voluntary exchange. Employers must use their resources to provide particular benefits in particular ways. Employees cannot exchange their labor for compensation that does not include the mandated benefits. Supposedly, mandated benefits are necessary to protect workers. In fact, mandated benefits harm workers by making them pay for benefits worth less than they cost.

Market incentives do far more than government mandates to motivate compensation packages that workers value most. Competition rewards employers who attract better workers at less cost. They do this by identifying benefits that their employees value more than the cost, and then substituting these benefits for monetary compensation. For example, assume that an employer can provide health insurance at a monthly cost of $200 per employee, and it is worth $300 to each employee. Providing the insurance and reducing the salary of each employee (or reducing the increase) by $250 per month makes the value of compensation to each worker higher by $50 and the cost to the employer lower by $50.

Employers are also in the best position to know what benefits their employees value most. Different firms provide different benefits because their employees have different preferences. Firms often provide a menu of benefits from which employees can choose, further targeting benefits to those valuing them most. The contrast with government mandates is considerable. Remote government authorities lack the local information needed to determine the appropriate mix of benefits in each firm, and have little incentive to make the best choices even if they did. This explains why mandated benefits tend to be one-size-fits-all propositions, with little adjustment to individual circumstance except in clumsy ways, such as exempting firms with fewer than some specified number of employees.

If a firm is not providing its workers with a fringe benefit, the chances are that workers value it by less than it costs. Obviously markets are not infallible. With worker preferences and technologies changing constantly, surely some fringe benefits worth more than they cost are not being provided. But it is a fantasy to hope government regulation is better than market incentives at providing the best mix of fringe benefits.

So why are mandated benefits so popular politically? The first place to look is for organized interests that benefit from them. While many believe that businesses oppose regulations like mandated benefits, economists argue that a host of regulations are means for businesses to protect themselves against competition. Often large firms in an industry lobby for a regulation with which they are already in compliance to increase the cost of smaller

competitors who are not. For example, Robert L. Crandall stated in 1987, when he was chairman and president of American Airlines:

> At American, we spend about $1,666 per employee per year – that's more than $80 million this year alone – on medical benefits for active employees and dependents.
>
> And we're spending $16 million a year for medical benefits for retirees … Yet Continental doesn't provide any medical benefits for retirees at all – and its active employees pay for most of their own health insurance. As a result, Continental's unit cost advantage vs. American's is enormous – and worse yet, is growing! [This is] why we're supporting Senator Edward Kennedy's legislation mandating minimum benefit levels for all employees. (Quoted in McKenzie 1988, p. 226)

Similarly, labor unions lobby for mandating benefits that their members already receive as a way of reducing competition from less-skilled workers. And those who would gain from supplying a fringe benefit have an obvious interest in having government mandate it. For example, chiropractors lobby to have their services mandated as part of employer-provided health insurance, and mental health professionals lobby to expand coverage to include a wide range of disorders they have labeled mental illnesses.[6]

Of course, organized interests such as small business organizations are harmed by mandated benefits, and they register their opposition politically. But the case for mandated benefits is easier to present in publicly appealing ways than the opposing case. For example, who wants to deny the advantages of family leave legislation to the parents of sick children, or of better health insurance for low-income workers? Advocates of mandating such benefits can emphasize their compassion and concern, while depicting their opponents as motivated only by the desire to save money. People feel more virtuous voting for mandated benefits, or politicians who favor them, than voting against them. As discussed earlier, because voting is largely an expressive activity with little personal cost attached, a small feeling of virtue from supporting mandated benefits can make a huge difference in its political prospects.

So lobbyists for mandated benefits have little trouble finding allies in legislatures, since politicians like benefits that are easily seen, that are widely recognized as laudable, and for which they can take credit. True, mandated benefits impose costs in excess of value by reducing reliance on private property and market exchange. But these costs are widely diffused, difficult to connect to the responsible legislation, and therefore easily ignored, and they would have little influence on the decisions of expressive voters even if fully recognized. So much of the costs of mandated benefits are ignored by politicians.

Restrictions on 'hostile' takeovers

The private ownership of the means of production is fundamental to free-market capitalism – indeed, this property right often serves as the definition of free-market capitalism. Since publicly held corporations are the dominant form of productive organizations in market economies, imposing restrictions on buying and selling privately held corporate shares undermines the advantages from market cooperation. Unfortunately, such restrictions are common.

The ability of people to buy and sell corporate shares at mutually agreeable prices is indispensable to the efficient allocation of capital. Corporations making better use of capital provide higher returns to their owners, as reflected in higher prices for their shares and lower cost of attracting capital. But free markets in privately owned shares do more than direct capital toward more efficiently managed corporations; they also serve to increase the efficiency in which corporations are managed.

It has long been recognized that the separation of ownership and control in publicly held corporations is a potentially serious problem.[7] The advantage of a firm's being able to attract capital from large numbers of shareholders means that the firm cannot be effectively managed by its many owners. Instead, the corporation has to be run by professional managers, and ideally they will put the shareholder/owners' interest in high profits ahead of their interest in management perks, privileges and job security. However, since a large number of shareholders cannot manage a corporation effectively, nor can they work together to ensure that managers do not put their interests ahead of the shareholders'. This is an example of the general problem of agency costs, a problem that can never be eliminated.[8] Corporate stock goes a long way, however, in reducing agency costs to reasonable levels, providing shareholders a significant degree of control over their managerial agents.

For example, the use of stock as managerial compensation can do a better job aligning the interests of managers with those of shareholders than a fixed salary. This is obviously incapable of providing a perfect alignment for a number of reasons, the most important being that top managers in large corporations receive only a small percentage of the corporate stock, even when generously compensated (see Jensen and Murphy 1990).

But the most important way that corporate stock imposes discipline on managers, at least for the purpose of my discussion, is through the market for corporate control. If the managers of a corporation are sacrificing profits to the pursuit of their own advantages, the malfeasance is reflected in the corporation's stock price – not perfectly of course, but there is no better barometer of managerial performance than prices of freely traded corporate shares. This means that information on managerial shirking is communicated in a way that motivates and allows corrective action. Depressed share prices due to poor management create an opportunity for someone to purchase a controlling inter-

est in the corporation, put in a better management team, and profit from the resulting increase in share prices.[9] For a variety of reasons, including the growth of unwieldy conglomerates, deregulation of the brokerage industry, and large cash flows into mature corporations, the market for corporate control became quite active from the late 1970s and well into the 1980s, with a significant increase in 'hostile' takeovers. There is compelling evidence that these takeovers created large wealth gains, as evidenced by the increased prices in the shares of firms taken over, not to mention the wealth-enhancing discipline takeover threats imposed on the managers of other firms.[10]

However, for the very reason 'hostile' takeovers create wealth – by imposing discipline on managers – managers strongly oppose them. The adjective 'hostile' for these takeovers reflects the influence managers have on the public debate. While many takeovers may be hostile from the perspective of managers whose jobs are at risk, they clearly are not hostile for owner/shareholders, who commonly receive premiums of 50 per cent and more for their stock. But it is easier for journalists to contact the few managers of a firm subject to a takeover threat than to get opinions from thousands of shareholders.

Corporate managers have no trouble enlisting politically influential allies opposed to takeovers. A firm's workers and, when relevant, their union representatives, see a takeover as a threat to their job security. Although in most 'hostile' takeovers a larger percentage of managers lose their jobs than do non-managerial workers, the efficiencies from takeovers often result from some job losses at many levels. The target firm often has plants in small communities that are likely to be downsized or shut down if a takeover succeeds, so political leaders in those communities will oppose the takeover. And chief executive officers (CEOs) of targeted firms are usually generous with shareholder profits, donating them to 'worthy' causes, many of them in the cities where they reside, cities anxious to show their appreciation to the CEO with lavish dinners, honors and political support opposing a takeover.

So there are real benefits from blocking takeovers. But the benefits from takeovers are invariably greater than the benefits from blocking them. But regulations that hamstring corporate 'raiders' create benefits that are very visible, highly concentrated, immediately realized, and easily credited to the anti-takeover regulations and the politicians sponsoring them. The benefits from corporate takeovers, though large in aggregate, are so diffused and delayed that they go almost completely unnoticed. They are spread over millions of consumers, shareholders and workers as slight gains for each, spread over a long period, with the connection between these benefits and opposition to anti-takeover regulations seldom recognized.

So the political process gives far greater weight to the costs of the market for corporate control than to the benefits. Even before the increase in take-

over activity in the 1980s, Congress had increased the cost of successfully taking over a firm with a tender offer, hostile or otherwise, with the 1968 passage of the Williams Act. But the biggest obstacles to takeover came in the 1980s from state anti-takeover regulations and state court decisions upholding them. These political actions sabotaged the market for corporate control, leaving incumbent CEOs, and their management teams, greater latitude to pursue self-serving objectives harmful to shareholders and the general public. Indeed, much of the malfeasance of corporate managers that has recently prompted calls for even more regulation is due to the protections given those managers by past regulation of the market for corporate control (see Manne 2002).

Global warming

I make no claims to knowing much about the science of global warming. How much, if any, the earth is warming, whether any warming that is occurring is a trend or the result of random variations in global weather patterns, or how much of that trend, if a warming trend exists, is due to human activity, are all questions I cannot answer. I am confident, however, that concern over global warming is being inflamed, inflated and used as an open-ended rationale for government regulation that diminishes the role of private property. I am also confident that a serious downside risk to this rush to regulate is going largely unnoticed.

Global warming is commonly described as one of humankind's gravest problems, threatening to flood coastal cities, spread virulent tropical diseases, increase violent storms, expand deserts and disrupt agricultural patterns, with the greatest harm being imposed on the poorest people – those living near the equator.[11] That's the bad news. The 'good' news is that since global warming is supposedly caused by humans, we can reverse its destructive effects by changing our behavior. Furthermore, we fortunately have experts who know the changes to make, so salvation is at hand if only we give them the necessary power and money.

There is a slight problem, however, with the good news: The experts recommend changes that require government controls, either directly or indirectly, over almost every aspect of our lives. Greenhouse gases, seen as causing global warming, are now being defined as pollutants that must be reduced significantly below current levels (as required, at least for developed nations, by the Kyoto Protocol). Carbon dioxide is receiving most attention, and reducing it would require tremendous lifestyle changes in the developed world (affecting everything from the type of products we consume to the type of occupations we pursue), and would make economic progress more difficult, especially in the less-developed world.[12] These changes could conceivably be made with the minimal degree of government regulations (though still a

lot) by establishing global markets for permits to emit greenhouse gases; however, the reality is that such markets would be heavily 'supplemented' with direct government controls and largely crippled by politically influential groups more interested in protecting their interests than in protecting the environment. Rather than relying on markets, political attempts to prevent global warming would mean government regulations being substituted for private property and market exchange on a massive scale.

I admit that without government action, market incentives would probably not reduce the emissions of greenhouse gases over the short run. But government regulations that hamper the market exchange of private property to reduce greenhouse gases would likely, even if successful, make any global warming problem worse.

There are two opposing approaches to global warming. The first which everyone talks about, I have already discussed – use government regulations to force greenhouse gas reductions. The other approach is to emphasize, not attempts to prevent global warming, but responding as efficiently as possible to any changes in global climate, whether from human activity or not. This latter approach would avoid government action that interfered with the superior ability of markets to provide the information and motivation necessary to adapt quickly and appropriately to changing conditions. While this approach *may not* do as much as direct government action to reduce global warming, it will result in better responses to any given increase (or decrease) in global temperatures.[13] So even if warming is greater under the 'market' approach than under the 'government' approach, the former might still be preferable. A more efficient response to a worse situation can be better than a less efficient response to a better situation.[14]

Even if the 'government approach' is more successful at reducing greenhouse gases than the 'market' approach, this success may have little, if any, effect on global temperatures, given the rather minor proportion of total carbon dioxide emissions caused by humans.[15] Furthermore, over the long run the innovation fostered by the disciplined freedom of the marketplace (recall the discussion earlier on) may offer the best hope for reducing reliance on fossil fuels responsible for most of the human release of greenhouse gases. By relying on market forces, with little thought to reducing greenhouse gases, rather than on government regulations specifically aimed at reducing them, we will probably do more over the long run to counter global warming (and almost everyone agrees that if global warming is a problem, it is a long-run problem), and do so by promoting economic prosperity and freedom instead of retarding them.

But with global warming, as with other issues, a strong political bias favors government regulation rather than markets. Most of the benefits from combating global warming (if successful) with government regulation will be

diffused and delayed, as will the benefits from market responses to any warming that occurs. So there might seem to be no bias favoring political responses resulting from immediate and concentrated benefits. But the politically salient considerations favoring government action are not the highly speculative benefits from preventing a small increase in global temperatures many decades into the future, but the immediate and concentrated benefits from larger bureaucratic budgets and research grants, and the political advantage of taking dramatic action directly targeted against a serious threat to humanity.[16] Certainly, bold and immediate action is much easier to sell to a frightened public than an argument for relying on private property and the market economy.

True, government regulations on greenhouse gases impose concentrated costs on business interests that are well organized politically. Clearly these interests have successfully prevented the US Senate from ratifying the Kyoto Protocol. But how successful will they be at opposing a series of small regulations, which in aggregate will seriously constrain the private sector, in the name of protecting the public against climate change? The case against such regulations is easily depicted as motivated by self-serving disregard for protecting the planet, which can activate considerable 'expressive voting' support for them. Furthermore, one should remember that, just as with mandated benefits, large companies often favor burdensome environmental regulation as a way of hampering competition from smaller rivals.[17] Certainly the excessive cost command-and-control environmental regulation has imposed on business (and the economy) has not prevented this approach from dominating environmental policy.

Conclusion

Governments tend to expand regulation far beyond the point where the marginal benefit of that expansion justifies the marginal cost from the erosion of the private property and market exchange upon which our freedom and prosperity depend. This tendency results from a distortion in the political process that exaggerates the benefits from government regulation while obscuring the benefits from the discipline of the marketplace. Interestingly, government regulation is politically popular for the very reasons it is economically destructive. Conversely, the marketplace is politically unpopular for the very reasons that it is economically productive.

Government regulation is popular because it can be used to provide protections and privileges to groups organized around a common interest. Because these benefits are immediate, concentrated on relatively few, and clearly connected to particular legislation, the politicians who support that legislation are enthusiastically appreciated by the beneficiaries, who show their gratitude with votes, favorable publicity and generous campaign contribu-

tions. The costs of the regulations from reduced economic efficiency are delayed, dispersed and not clearly connected to the regulation that caused them. Therefore, the costs of regulations are largely ignored by the public, so that politicians can impose them with a large measure of impunity. This not only makes regulations politically popular, it also means that they are invariably expanded to economically destructive levels.

Markets are economically successful because they concentrate costs in the form of layoffs, bankruptcies and loss in asset values on those not using their resources to best respond to the interests of others. These costs, and the adjustments they motivate, are painful, and those suffering from them naturally blame the harsh competition of the marketplace. On the other hand, the tremendous wealth created by markets is so widely distributed that people tend to take it for granted as part of the natural order of things, not realizing that it depends on the hardships from market competition – hardships which seem all the more unfair and objectionable because of the general wealth of society. One is reminded of Machiavelli's ([1513] 1988, p. 60) comment about those who 'admire [the] achievement … yet condemn the main reason for it'.[18]

So market failure is commonly attributed to the very things that make it is so successful at creating wealth and fostering freedom. Advocates of government regulation as the answer to these market 'failures' are either organized interests knowingly using government to capture private benefits at public expense, or people being duped by the public-interest rhetoric of those interests. In either case, the political pressure is for more regulation that sabotages the private property and market exchange that is the source of our prosperity and freedom.

Notes

1. Friedman (1962, p. 18) states, 'The [private] suppliers of paper are as willing to sell to the *Daily Worker* as to *The Wall Street Journal*. In a socialist society … the hypothetical supporter of capitalism would have to persuade a government factory making paper to sell to him, the government printing press to print his pamphlets, a government post office to distribute them among the people … and so on'.
2. The best policy for reducing most forms of pollution would be to issue transferable pollution permits as private property, and let those permits, and the pollution they allow, be allocated in response to the market prices that would emerge for them. This would not be a perfect market, since the supply of these permits would be determined politically instead of through markets, and so the permit prices would not likely reflect the marginal cost of pollution very closely. But such markets would do a better job motivating least-cost pollution reduction than the command-and-control policies that currently dominate pollution policy.
3. Also, even though steady productivity increases yield dramatic benefits over time, slow but steady improvements are easily ignored. Cox and Alm (1999, pp. 39–51) point out how almost everything we buy has become less costly over time, where the cost of an item is measured, as it should be, in terms of the hours of work necessary to earn the money to buy it. For example, in 1950 the average worker had to work two and a half hours to earn

enough to pay for a three-minute long-distance call. Today the same three-minute call costs only two minutes of work time. And this does not take into consideration that many of the products we buy were unavailable at any cost a few years ago, that even when a product was available in the past it is much better today, and that the fewer hours of work necessary to earn the money for today's products is spent doing work that is more pleasant and safer than in the past. Yet, people constantly complain about the high cost of living.

4. Workers will not necessarily make more in their new jobs than in their old ones, though many will. Their compensation in their old jobs reflected the value they produced in those jobs before changes rendered them less valuable relative to the value they can add in new jobs. So even when displaced workers do not experience increased compensation in their new jobs, they still receive more in those jobs than they would make if they remained in their old ones and were paid what those jobs are currently worth.
5. Brennan and Lomasky (1993) provide the most detailed discussion of the cause and implications of this expressive voting.
6. For example, the Mental Health Liaison Group, a self-described coalition of over 50 national mental health advocacy groups, is lobbying the US Congress to amend the Mental Health Parity Act of 1996 to expand mental health benefit limits under private health insurance coverage. See the website at www.mhlg.org. for more on the Mental Health Liaison Group's lobbying activities.
7. In a publicly held corporation, the shares are privately held by members of the general public, instead of privately held by an individual or small group.
8. See Jensen and Meckling (1976) for the seminal article on agency costs in corporations.
9. See Manne (1965) for the first systematic treatment of this market for corporate control.
10. For a summary of the studies on the wealth effect of corporate takeovers, see Jarrell et al. (1988).
11. According to former Senator George Mitchell, in his book *World on Fire: Saving an Endangered Earth*, left unchecked, global warming 'would trigger meteorological chaos – raging hurricanes … capable of killing millions of people …record-breaking heat waves; and profound drought that could drive Africa and the entire Indian subcontinent over the edge into mass starvation. … Unchecked, it [global warming] would match nuclear war in its potential for devastation' (quoted in Moore 1995, p. 83). Moore points out that the earth has experienced periods of weather noticeably warmer than currently, some of them relatively recently (the last few hundred years), and that they have been associated with bursts of human progress and improvements in living standards, whereas periods of cooler weather have been periods of stagnation, and worse.
12. Human role in carbon dioxide discharges is modest compared to nature's role. According to Easterbrook (1995, p. 312), 'naturally occurring carbon emissions outnumber human-caused emissions roughly 29 to one'. Interestingly, some scientists believe that methane may contribute as much to global warming as carbon dioxide because, though less prevalent, it is far more effective at trapping heat. And it would be much less costly to reduce methane. See Easterbrook (ibid., pp. 298–300) for the advantages of focusing on methane, and some of the special-interest opposition to doing so.
13. I emphasize *may not* because, as explained below, generalized market forces may actually reduce greenhouse gases indirectly by more in the long run than will government regulations designed to do so directly.
14. In this regard we might consider seriously Nordhaus's (1993, p. 23) observation, that 'perhaps we should conclude that the major concern lies in the uncertainties and imponderables impacts of climate change rather than in the smooth changes foreseen by the global models'. It should be pointed out that Nordhaus uses this observation to emphasize the importance of flexible policy approaches rather than to recommend market adjustments to unforeseen conditions.
15. This argument is stronger for carbon dioxide than for methane emissions, which are more easily reduced and may be as responsible for global warming. See note 12.
16. If these benefits were not important in global-warming politics, it would be difficult to explain the frightening scenarios those who benefit from political action against global warming are constantly putting before the public. For example, in a fit of candor, Stephen

Schneider, a major activist in the fight against global warming (who in the 1970s was warning of the coming ice age) told the *Boston Globe* in the early 1990s, 'It is journalistically irresponsible to present both sides (of the global warming issue) as though it were a question of balance… I don't set very much store by looking at the direct evidence. … To avert the risk we need to get some broad-based support, to capture public imagination. That, of course, means getting loads of media coverage. So we have to offer up some scary scenarios, make some simplified dramatic statements and little mention of any doubts one might have. … Each of us to decide what the right balance is between being effective and being honest' (quoted in Bandow 1998, p. 35).

17. See Malone and McCormick (1982) and Parshigian (1984) for two important articles on business support for inefficient environmental regulations.
18. The harshness and cruelty of Hannibal's command of his army was the reason for the admirable achievement discussed by Machiavelli. As with the harshness, or 'cruelty', of the market, no one wants to suffer the discipline imposed, but everyone is better off when that discipline is imposed on others. Justice requires that the discipline be imposed on all who benefit from it – allowing some to avoid that discipline while benefiting from its imposition on others would be unjust.

References

Bandow, D. (1998), 'Global politics, political warming', *The Freeman*, **48** (1), January, 34–6.

Brennan, Geoffrey and Loren Lomasky (1993), *Democracy and Decision: The Pure Theory of Electoral Preference*, Cambridge: Cambridge University Press.

Cox, W. Michael and Richard Alm (1999), *Myths of Rich and Poor*, New York: Basic Books.

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

Easterbrook, Gregg (1995), *A Moment on the Earth: The Coming Age of Environmental Optimism*, New York: Penguin Books.

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

Gwartney, James and Robert Lawson (2002), *Economic Freedom of the World: 2002 Annual Report*, Vancouver: Fraser Institute.

Hayek, Friedrich A. (1960), *The Constitution of Liberty*, Chicago: University of Chicago Press.

Jarrell, G.A., J.A. Brickley and J.M. Netter (1988), 'The market for corporate control: the empirical evidence since 1980', *Journal of Economic Perspectives*, **2** (1), Winter, 49–68.

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.

Jensen, M.C. and K.J. Murphy (1990), 'CEO incentives – it's not how much you pay, but how', *Harvard Business Review*, May–June, 138–53.

Machiavelli, Niccolò (1513), *The Prince*, edited by Quentin Skinner and Russell Price (1988), Cambridge: Cambridge University Press.

Malone, M.T. and R.E. McCormick (1982), 'A positive theory of environmental quality regulation', *Journal of Law and Economics*, **25**, 99–123.

Manne, H. (1965), 'Mergers and the market for corporate control', *Journal of Political Economy*, **73**, April, 110–20.

Manne, H. (2002), 'Bring back the hostile takeover', *The Wall Street Journal*, 26 June, A16.

McKenzie, Richard B. (1988), *The American Job Machine*, New York: Universe Books.

Moore, Th. G. (1995), 'Why global warming would be good for you', *Public Interest*, **118**, Winter, 83–99.

Nordhaus, W.D. (1993), 'Reflections on the economics of climate change', *Journal of Economic Perspectives* , **7** (4), Fall, 11–25.

O'Driscoll, Jr., Gerald P., Kim R. Holmes and Mary Anastasia O'Grady (2002), *2002 Index of Economic Freedom*, Washington, DC: The Heritage Foundation and Dow Jones & Co.

Parshigian, P.B. (1984), 'The effects of environmental regulation on optimal plant and factor shares', *Journal of Law and Economics*, **27**, 1–28.

Read, L.E. (1958), 'I, pencil', *The Freeman*, December, 32–7.

Smith, Adam (1776), *An Inquiry into the Nature and Causes of the Wealth of Nations*, vol. I, reprinted (1981) in Indianapolis, IN: Liberty Fund, Inc.

15 Corruption

*Bruce L. Benson and Fred S. McChesney**

I won't tell anybody else
I'll keep it to myself
So c'mon and steal away.
Please, steal away.
(Jimmy Hughes, 'Steal Away'[1])

Introduction

Corruption is ubiquitous, not just in the world but also in the literature of law and social science. 'The literature on corruption is both vast and diverse' (Bowles 2000, p. 462). Very little of that literature, however, views corruption in property rights terms. Yet, it is submitted here, the problems of corruption are essentially those related to property rights. It follows that desirable solutions to corruption problems must start by recognizing the property rights nature of the phenomenon.

Broadly defined, corruption is not limited to the public sector. For example, an employee of XYZ, Inc. may take a kick-back from Firm A in exchange for the employee's awarding a contract from XYZ to A rather than to Firm B, a lower bidder. Private sector corruption, however, is a species of agency cost more generally, and thus raises issues beyond the scope of this chapter. The subject here is corruption in the public sector.

A summary such as this can only outline the principal modes of analysis to which corruption has been and should be subjected. Some more minor analytic strains must perforce be shortchanged. The works referenced below contain many citations to other work not discussed here. Fiorentini and Zamagni (1999) compile a good bit of the economic scholarship (see also Kaufman 1998).

The standard view of corruption

The standard economic discussion of corruption has been more journalistically descriptive (with background whiffs of sociology) than economically analytic. It has also relied more on law than on economics to define and evaluate corruption. Although economists' presentations typically blur the distinction between normative and positive evaluation, it is useful to separate the two.

Normative evaluation of corruption

First, economists have been confused about what 'corruption' means. Definitions like 'an illegal payment to a public agent to obtain a benefit' for a private individual or firm (Rose-Ackerman 1999, p. 517), or 'the illegitimate use of public roles and resources for private benefit' (Bicchieri and Duffy 1997, p. 61) are prevalent. Corruption then is condemned out of hand because it is illegal or illegitimate, terms that are not economic, just pejorative. 'Most scholarship assumes that corruption is bad and focuses on alternative ways to attack it' (Mookherjee and Png 1995, p. 145). Prominent examples include Alam (1990), Mauro (1995, 1998), Ades and Di Tella (1997), Klitgaard (1998), Tanzi and Davoodi (1998) and Rose-Ackerman (1999).

For instance, Abbott and Snidal (2002, pp. 158–60) explain how the World Bank formerly regarded payments to government officials as an inevitable necessity for cutting through red tape in order to implement development policies. Pressure mounted for the World Bank to demand action against corruption, however, and when James Wolfensohn became Bank President in September 1996, he declared that the Bank would do everything that it could to battle the 'cancer of corruption' (quoted in ibid., p. 159). The World Bank implemented new and extensive anti-corruption policies the next year.

This view that all corruption is bad now appears to dominate analysis. As Bowles (2000, p. 475) reported, 'The conventional wisdom is that corruption is harmful, and for some authors this seems almost self-evidently to be the case'. (Bardhan 1997 presents a minority, dissenting view; see also Abbott and Snidal 2002.)

This approach raises two problems. First, statically, it is not always clear what constitutes an 'illegal' or 'illegitimate' payment. The same authors who define corruption as payments to public agents for private benefit also note that, by local custom, 'gifts' to public officials are not popularly viewed as corruption, whatever their legal status (Bowles 2000, p. 462). Indeed, concessions to local custom and practice are urged. 'Payoffs that are widely viewed as acceptable should be legalized' (Rose-Ackerman 1999, p. 519). Alternatively, the term 'extra-legal' is used to refer to payments that are nominally illegal but tolerated (Leff 1964).

Second, the approach of defining corruption in legal terms (with some slack cut for local definitions of 'gifts') effectively subordinates economic analysis of a practice to legal definitions. As legislators alter the law, what constitutes corruption necessarily changes as well. In the United States, for example, laws concerning payments to political candidates change frequently, in the name of preventing corruption.

Defining corruption legally, however, means that a change in the law converts something that was economically unobjectionable (or even benign) to something that is, by definition, economically malign. What was legal

today becomes illegal (corrupt) tomorrow. Particularly given the economist's general suspicion that much legislation is motivated by something other than legislators' public-interest desire to improve mankind, condemnation of a practice because politicians choose to validate or invalidate it sells economic science short.

An alternative normative approach equates corruption, not with illegality, but with immorality. Bowles (2000, p. 462) summarizes this approach: 'In everyday use corruption is a term which conveys an element of moral disapproval. It represents an unwelcome deviation from some desired state of the world. … It is implicit in most discussions of this kind that elimination of corrupt transactions would self-evidently improve welfare'. But, perhaps not surprisingly, the moral approach has yielded little economically compelling insight. (See Bicchieri and Duffy 1997, p. 61 and n. 3 for further discussion.) Economics has no more to contribute to morality-based condemnations of corruption than it adds to lawyers' definition of what is corrupt.

Nevertheless, the standard accounts have offered a few economics-based reasons against corruption, the most cogent falling into two categories. First, corruption entails Tullock (1967) costs in the dead-weight loss caused by competition for corrupt payments (Rose-Ackerman 1999, pp. 2, 147, 213; Bowles 2000, pp. 464–5). In effect, corruption is seen as a species of rent seeking, a point developed further below.

Second, some analysts posit, corruption actually can make the public sector work better. 'Outside' (extra-legal) payments may elicit superior bureaucratic service, and so might be compared to restaurant tipping. Among the 'possible benefits of corruption', Bicchieri and Duffy (1997, pp. 61–2) mention that 'it can speed up cumbersome procedures, bypass inefficient regulations, buy political access for the excluded, thus fostering the integration of immigrant or parochial groups, and even produce policies that are more effective than those emerging from legitimate channels'. This normative position has long been argued (for example, Leff 1964) but not generally accepted in economist circles: 'it should be stressed that this is rather a minority view' (Bowles 2000, p. 475).

Positive analysis of corruption

Although normative writing on corruption has dominated the traditional literature, positive analysis has begun to emerge. The positive literature has largely been of two sorts. There have been a few attempts to assemble positive microeconomic models of corruption, but very little empirical testing of their implications. There are many more attempts to measure the macroeconomic effects of corruption.

Traditional positive models of corruption The law makes much corruption a crime and economists define corruption as payments that are illegal. So, standard economic models of crime (for example, Becker 1968) often furnish the basis for positive analyses of corruption. The crime-based models of corruption naturally parallel those for crime generally. For example, corruption can be fought with carrots or sticks. Increasing fines for corruption will decrease the predicted incidence of corruption. Increasing the wages of civil servants in effect raises the penalty to them of being detected and fired for corruption, and so should also decrease the amount of corruption (see, for instance, Rose-Ackerman 1999, pp. 71–5, Bowles 2000, pp. 467–8).

The applicability of the Becker model is limited, however. Some corruption is analogous to crimes like theft. But much corruption is more like a contract, making both sides better off, rather than the typical crime, which benefits one side at the expense of the other (Bowles 2000, pp. 464–6). The implications of this distinction have not been fully developed or even appreciated, as explained below.

Though typically treating corruption as a subset of crime generally, the corruption literature has nonetheless developed particular themes. Corruption may be a 'lemons' market: once some corruption takes hold, it may (under certain assumptions) take over (for example, Rose-Ackerman 1999, p. 16; Acemoglu and Verdier 2000). Then again, it may ebb and flow (for example, Bicchieri and Duffy 1997). Imposing penalties for corruption on corrupt bureaucrats may reduce the amount of corruption, but then again it may just raise the price of corruption, that is, the amount paid for a corrupt act (Mookherjee and Png 1995). Or it may be more complicated than all that (Basu et al. 1992).

Empirical analysis of corruption In short, positive theoretical models of corruption reach very different conclusions, with contrasting empirical implications. Furthermore, none of these theories has been either validated or refuted empirically; there have been no systematic efforts to test the various crime-based microeconomic models of corruption. As Ades and Di Tella (1997a, p. 496) note, theories abound but 'a lack of data on corruption to test the theoretical contributions [has] allowed conflicting theories on the causes and consequences to coexist'. Lack of testing is typically ascribed to the clandestine nature of the activity: 'one of the central features of corruption is that it is, in essence, a cooperative venture which both participants have every incentive to keep secret. It is thus extremely difficult to identify either particular instances of it or its extent in aggregate' (Bowles 2000, p. 467).

This argument is unconvincing. *All* crime is something its perpetrators would like to keep secret, and yet, empirical studies of many sorts of crime are legion. Rather, the lack of micro-level testing of various hypotheses

concerning corruption seems due to the rarity with which corrupt officials are apprehended and punished. That is, the problem is not the impossibility of getting data – Svensson (2003) shows that it can be obtained – but rather a lack of interest politically in pursuing corruption in the first place.[2] The low incidence of detection and prosecution of corruption is discussed below.

Lacking micro-level data, empirically-minded economists instead have focused on corruption's macroeconomic effects, using national indices for corruption developed by research organizations. (Data sources are described in Heidenheimer, 1996 and in Ades and Di Tella 1997a) Bardhan (1997, pp. 1327–30) provides a useful summary of and citations to the literature on corruption and macroeconomic growth; for a particular example, see Ades and Di Tella (1997a). A related literature has examined the macro factors that might explain the cross-national incidence of corruption (Bardhan 1997, pp. 1330–34).

Corruption in the Third World Although analytically equivalent to corruption generally, corruption in Third World (or underdeveloped) economies has been a particular focus of corruption analysts (for example, de Soto 1989; Kaufmann 1998; Cho and Kim 2001). A principal problem identified by many scholars has been the way that corruption interferes with the proper functioning of government. In developing countries especially, corruption severely impedes collection of taxes, enforcement of regulations, and management of public sector enterprises (Mookherjee and Png 1995, p. 145).

If the problem of corruption for the Third World is interference with governmental functions, it might seem to follow that privatization of at least some of those functions would be an improvement. However, privatization has not been the solution ordinarily proposed. Perhaps surprisingly, shifting resources and transactions from the public to the private sector has been seen as increasing the potential for corruption, and therefore to be discouraged, not applauded. Rose-Ackerman (1999) complains that a Third World dictator might privatize by transferring resources to himself or his family. Even routine privatization offers corruption opportunities to bureaucrats responsible for privatizing, including insider trading, taking bribes and selling tax breaks. Thus, in the ordinary analysis privatization merely alters the type but not the amount of corruption.

The property rights view of corruption

From the foregoing, it seems that a good deal of economists' writing about corruption is not economic at all. Corruption is typically defined by what is illegal, and automatically deplored. So, as the law shifts the boundaries of corruption, economists automatically shift their normative evaluation of any given practice. More fundamentally, the traditional economic perspective on

corruption forsakes the accepted economic approach to human activity. There is no underlying model leading to the conclusion that corruption is bad, only an assumption to that effect.

The appropriate model is one based on government actors' property rights. That is, a useful economic model must start with what economists know about property rights. And then, because the property rights at issue belong to government officials, the property rights points must be combined with elements of public choice economics.

The public official as homo economicus

Because government officials are rational maximizers of their own welfare, economic analysis of public corruption begins with the opportunities available to government actors to benefit themselves. Political benefits are of different sorts. To the extent that their position as elected officials depends on winning votes, public officials will (all other things equal) choose policies that will generate electoral reaction sufficient to keep them in office. Similarly, to the extent that their position as bureaucrats depends on satisfying the elected officials who vote on their budgets, officials will operate bureaucratically so as to maintain themselves in office. Tailoring decisions to satisfy the electorate or elected officials, however, is the essence of democracy, and so does not ordinarily figure in analyses of corruption.

A property rights approach to corruption focuses on how public officials gain from their offices. It has nothing to do with whether officials' wealth increases are legal or moral. Nor does defining corruption as actions increasing public officials' personal wealth imply any normative evaluation. It is an entirely positive description, one which will be seen to entail varied normative conclusions economically. It makes clear, however, that the source of corruption is government officials' property rights. Once the property rights perspective to corruption is understood, the idea that corruption is something that happens despite, rather than because of, politicians and bureaucrats is no longer tenable. Leff (1964, p. 10) makes the point in the context of underdeveloped economies:

> The critique of bureaucratic corruption often seems to have in mind a picture in which the government and civil service of underdeveloped countries are working intelligently and actively to promote economic development, only to be thwarted by the efforts of grafters. Once the validity of this interpretation is disputed, the effects of corruption must be reevaluated.

If corruption is defined as use of one's public office to accrue private wealth, the issue then becomes how officials can benefit themselves through their office. In property rights terms, two questions are paramount. First, what does the public official own (Cheung 1996)?

The government official's principal possession is power. Government is, by definition, the monopoly owner of the right to coerce the citizenry. As Friedman (1973, p. 152) writes, 'The special characteristic that distinguishes government from other agencies of coercion (such as ordinary criminal gangs) is that most people accept government coercion as normal and proper'. Part of government's power (and thus ability) to coerce is the monopoly right to assemble armies and police forces, maintain prisons and levy taxes – none of which is possible for private individuals.

Bardhan (1997, p. 1321) adds a critical point when he defines corruption as 'malfeasance for private enrichment which is difficult to monitor'. Difficulty of monitoring public officials' behavior allows them to enrich themselves by discretionary use of the power they possess. In the absence of monitoring costs, agents (for example, legislators, bureaucrats) will generally have to serve the interests of their principals (voters in a democracy; the aristocracy, the dictator, the official party, the mafioso, or other principals in other power arrangements). Further analysis of monitoring costs follows below.

If status as a government official confers a property right in power over private citizens, the second question is presented. How can officials use that discretionary power to benefit themselves?

Forms of corruption

Generically, there are three ways in which politicians can benefit themselves through the power in their office. Presumably, as rational maximizers, politicians will equate at the margin the returns from each of the three means, so as to maximize their return overall. Thus, 'corruption' in the ordinary situation (region, country) will entail a mix of each of the three forms of corruption. Each of the three, however, arises in a different institutional setting, with different positive and normative implications economically.

Taking The most obvious way that the more powerful can benefit themselves is by just taking from the less powerful, and so, outright theft (or embezzlement) is one common form of corruption; that is, of the use of governmental position to alter private property rights. This is rightly condemned by students of corruption (see Rose-Ackerman 1999, pp. 114–21). The standard economic analysis of ordinary theft (Tullock 1967) applies just as well when the thief is a government official, the legislature or even the sovereign himself.

Corruption by theft is perhaps most often associated with the Third World, where 'kleptocrats' (to use Rose-Ackerman's apt phrase) are infamous. But corrupt politicians and public officials are commonplace in more advanced economies as well. Corrupt bureaucrats are likely to be in a position to discriminate among potential buyers, refusing to sell to those that, for some

reason, are 'less desirable', unless they are paid. For example, a policeman may give a traffic ticket to someone who does not show sufficient 'respect' (or is black, or Hispanic, or female, or unattractive), but not to someone who is polite (or white, or male, or attractive), unless the policeman is paid. When the power to take is present, it predictably will be used.

It is important to note, however, that not all governmental takings may be illegal. Nationalizations, for example, are economically deleterious, but perfectly legal, since what is legal is defined by the nationalizing government itself. Likewise, legislatures define what is embezzlement. From a corruption perspective, there is no important difference between a despot stealing (illegally) from the treasury to build himself a palace and the legislature (legally) taxing citizens to obtain money to build the despot a palace.

Rent seeking/bribery A second way politicians can benefit themselves by use of their power is taking bribes in exchange for exercise of their power in ways that benefit the private payer. Bribery is a form of corruption routinely decried by economists, although the distinction between bribery and gifts 'is a cultural matter' (see Rose-Ackerman 1999, p. 110). Since gifts are legal, the normative evaluation of any particular exchange is indeterminate in the standard analysis.

What some call 'gifts' economists call 'rent seeking'. The use of government to obtain special favors is now a well-understood phenomenon (Tullock 1967, 1993). Corruption is often characterized as a type of 'rent seeking' (Rose-Ackerman 1999, p. 2), but more fundamentally it is a property rights problem. If property rights were clearly delineated and perfectly secure there could be no rent seeking. Rent seeking arises because the discretionary power of the state to alter or transfer property rights through taxation and spending, regulation and eminent domain makes property rights insecure. Likewise, licensing or franchising, price ceilings, zoning, tariffs and quotas, job requirements, and other similar acts attenuate the rights of some individuals to use their property in ways that they desire – and all are sources of potential corruption. Both theft and rent seeking arise because property rights are not perfectly and completely delineated.

From this perspective, corruption is not a 'type' of rent seeking. It is a part of the rent-seeking process. Corruption occurs when an individual or organization has power to influence the allocation of private property rights to assets. When that individual or organization is part of the apparatus or the state (for example, with legislative or enforcement powers), it might be said that some of the rights to the asset belong to the 'public', but in fact, they are controlled by legislators, judges or bureaucrats with actual decision-making powers.

Of course, the favors do not come free. However, the exchange – compensation paid to politicians for private rents – is frequently quite legal. If

properly and skillfully arranged, payments ('gifts') to politicians will not be treated like illegal bribery. But economically, the two are essentially equivalent, both positively and normatively. The analysis of payments from rent seekers to public officials is virtually identical whether a payment is called a bribe, extortion, a gift, or a campaign contribution. All such payments are 'corrupting' if they alter (corrupt) the public official's decision relative to what it would be in the absence of the payment by inducing an allocation of property rights that allows the payer to collect rents that would not otherwise be available. There is growing evidence that 'legal' payments such as campaign contributions do in fact alter political decisions (Stratmann 1991, 1992, 1996, 1998). Hence, whether legal or illegal, such payments are examples of corruption.

Rent-seeking legislation generally reflects the demands of influential interest groups. The cost of organizing most groups may be high. Once a group has been organized, however, the marginal cost of demanding more property rights alterations (more bribery or rent seeking) is low. Organized groups therefore are likely to demand subsequent rights alterations that generate far fewer benefits than would have been sufficient to induce them to organize in the first place.

Thus, a property rights perspective provides an economic model for the claim that 'corruption breeds more corruption' (Rose-Ackerman 1999, p. 3). As organized groups demand more changes in property rights, those whose rights are threatened have increasing incentives themselves to organize for protection. But once organized, they too can demand changes in other property rights that threaten still more people. Thus, a spiraling process of more and more rent seeking develops (Buchanan and Tullock 1962; Olson 1965; Benson 2002), making property rights increasingly insecure (Benson 1984).

Others have noted the bribery/rent-seeking spiral. Peruvian governments have passed about 28 000 laws and regulations per year since the Second World War. De Soto (1989) assessed the impact of this legal morass on the citizens and economy of Peru by seeking the legal approvals and licenses necessary to set up a two-sewing-machine garment factory in a Lima shanty town. Working with five university students to navigate the bureaucratic maze, De Soto found that establishing such a small business legally took 289 days and cost 31 times the average monthly minimum wage. They were asked for bribes ten times, and ultimately had to pay twice in order to proceed. Subsequently, de Soto (2000) has described the same sort of legal and bureaucratic quagmires in several other countries, and the same results arise: large informal underground sectors and corruption.

The bribery/rent-seeking corruption spiral has larger implications for analysis of legal systems. Rent seeking leads to too many laws.[3] The laws are increasingly deleterious economically. The legislative process and the rents

that it generates can be seen as a common pool. As with any common-pool resource, it is overused, leading to crowding and rapid decrease in the 'quality' of the output. The excess use of a common pool can be offset by investments in maintenance and improvements. But because the individuals who make such investments cannot exclude others from capturing the benefits, they have virtually no incentives to make the investments.

As Demsetz (1967) stressed, property rights are only created if the benefits of doing so exceed the costs. The fact is that the typical voter–taxpayer does not really know what is being purchased with tax revenues or what wealth is being transferred and destroyed through bribery/rent seeking. The lack of knowledge about government is perfectly rational, as citizens have very weak incentives to obtain the information required to effectively evaluate government performance, even in a democracy. After all, there is no guarantee that the individual's evaluations, after information is obtained, will matter at all.[4]

Even interest groups who benefit by illegal bribery or by rent seeking likely will have little information about the costs of the rents they purchase. Rents are benefits transferred from others. The beneficiaries do not want information about those costs (or the magnitude of their benefits) available, lest those paying the costs will find out. Likewise, bureaucrats capturing personal benefits by collecting bribes or creating rents do not want the costs of their activities known (Benson 1985). Breton and Wintrobe (1982, p. 39) emphasize that bureaucrats selectively release both true and false information to the legislature, the press, the public, the interest groups benefiting from their activities, and those that might be opposed to them. This 'selective distortion' means that the costs of effective monitoring to any of these groups will be quite high, so not much time or effort is expended in monitoring the cost-effectiveness of bureaucratic performance either.[5] And so bureaucrats generally have a great deal of discretion, creating an environment in which corruption can flourish (Benson 1981, 1988a, 1988b; Benson and Baden 1985).

High monitoring costs are not the only reason to expect bureaucratic corruption, however. With excess legislation, these agencies cannot possibly produce all of the rent-transferring goods and services mandated or enforce all of the transfer rules that are created. So, decisions must be made regarding how to ration the limited resources that bureaucrats control among the competing demands of individuals and interest groups. If no rationing institution is established, rationing by first-come–first-serve rules often evolves. But those rules lead to congestion, so the service is actually rationed by waiting. Those wanting the product rush to line up early and wait to be served, and only those willing to wait can get the product. Rationing by waiting is very prevalent in the public sector.[6] In other words, crowding or congestion occurs, as with any common pool.[7]

Corrupt bureaucrats can discriminate by giving benefits to those who pay them (Svensson 2003). Black markets inevitably arise when a good or service is rationed by waiting, so that markets clear only with illegal sales. Black markets arise in the provision of bureaucratically produced rent transfers (property rights alterations), as bureaucrats take bribes in exchange for the 'right' to get moved up in the queue, or to avoid the enforcement of a particular rule in order to do something illegal (Benson 1981). A bribe accepted to not write a traffic ticket when the driver really has been speeding is, essentially, the illegal sale of the right to exceed the posted speed limit. Similarly, police can and do accept payments to leave a gambling operation or a drug dealer alone, and perhaps to harass any competition that the gambler or dealer may have, essentially selling the right to operate in the illegal market as a monopolist because police are selectively excluding their competition (Benson 1981, 1988a, 1988b, 1990, pp. 159–75; Benson and Baden 1985; Thornton 1991; Rasmussen and Benson 1994, pp. 107–18).

Rent extraction/extortion Rent extraction is a third strategy by which public officials can increase their personal wealth. The term refers to public officials' ability to threaten private parties' personal wealth, but then be paid not to make good on their threats (McChesney 1987, 1997). For example, the Clinton administration threats to impose price controls on the health-care industry (proposed in 1993 but abandoned the next year) generated torrents of private money 'contributed' to politicians, who were paid *not* to legislate. 'Your money or your life' is the street-crime equivalent of political rent extraction: private parties pay government officials rather than lose something of even greater value.

Rent extraction is to rent seeking as torts are to contracts. With rent seeking, private and public parties have a deal making both parties better off (although others who are not parties to the bargain are worse off), in a contract that constitutes legal bribery. The politician gets the money, the payer gets the favor. Rent extraction, however, is legal extortion. The private citizen pays, not to be better off, but to avoid being made worse off. Consider the corrupt payments to Nazi officials by Jewish freedom fighters in the Warsaw ghetto during the Second World War to smuggle arms into the ghetto and spirit Jews out.

Normatively, rent extraction differs from rent seeking. To return to the Clinton health-care threats, price controls are costly to society. So, paying off legislators not to impose price controls is beneficial. Admittedly, the benefits accrue from a second-best situation; in a first-best world, the government would not possess the discretionary power to impose price controls in the first place. But given a second-best (that is, the actual) world, rent extraction increases social wealth. As Huntington (1968, p. 69) writes, 'the only thing

worse than a society with a rigid, over-centralized, dishonest bureaucracy is one with a rigid, over-centralized, honest bureaucracy'.

Much rent-extracting corruption is legal. Payments made to avoid Clinton-administration price controls were entirely legal. The payments from those trapped in the Warsaw ghetto were illegal. But as these contrasting examples show, the fact of legality or illegality tells one little about the economic desirability of the practice. Whether legal or illegal, the payments achieved a desirable goal: freedom from economically undesirable governmental power.

Corruption and government

Government size and corruption

From the foregoing, two important conclusions about the role of government in corruption emerge. First, it is unimportant allocatively how the law (itself a product of government) defines corruption. Some outright taking by a despot or kleptocrat may be legal; some of it may be embezzlement. But economically, it is theft, with all the attendant Tullock (1967) costs. Some payments for special favors may be legal rent seeking; others may be bribery. But the economic losses are again those that Tullock and others (for example, Krueger 1974) have identified. Protection money paid to politicians to avoid threatened losses may be legal campaign contributions; it may be illegal extortion. Regardless, however, the losses are those indicated in the rent-extraction model.

Thus, the true importance of making some corruption legal and other corruption illegal is distributional rather than allocative. Outlawing a particular practice legislatively (for example, prostitution or drugs) creates the opportunity for non-legislators (bureaucrats, including police) to sell illegal relief from the legislative prohibition, and so to profit personally. Legislation that merely regulates an activity allows legislators themselves to sell regulatory services that legally benefit one group at the expense of another. An interesting point, not considered here, is why legislators would ever make something outright illegal, as that effectively increases the value of the property rights held by lower-level government officials.

Moreover, the source of all three sorts of corruption is government power over private property rights. It follows that corruption is a positive function of government size, a proposition validated empirically by Goel and Nelson (1998) using American state data. Holding constant other variables, they find (1998, p. 116) that 'the number of cases of actual abuse of public office increase disproportionately with the size of state-local government'.

Or, consider Russia. Simons (1996, pp. 251, 254) notes that 'the system is changed but state control is not', in part because of 'the mere tradition of

state control (especially as concerned the foreign, non-state sector in the Soviet Union)'. While there have been some very visible reductions of state control in some areas, such as privatization of state enterprises,

> the state has to date seemingly been unable to adequately define its role [regarding] … the important requirement … of market(-type) economies that parties enjoy basic freedom of contract … Rather, the role of the state at present is as yet an ill-defined one and where it takes shape it often appears to favor too much control when less is required … And it is precisely this ill-defined role of the state – not as the owner of all the 'means of production' as was previously the case in the Soviet era but as a regulator of economic activity in the transition period – which is a serious impediment to significant progress in the transition process. (Ibid., pp. 256–7)

This vast array of complex rules enforced by bureaucrats with a great deal of discretionary power creates a lucrative situation for public officials who are willing to accept bribes. As Dempsey and Lukas (1998, p. 471) suggest, 'The Russian government's reluctance to give up further economic control is the single greatest catalyst to organized crime'.[8] In Russia, much of organized crime is not involved in the drug, prostitution, or gambling markets that are the primary focus of such groups in the United States and Western Europe. Instead, many 'mafia' are serving as middlemen to facilitate the illegal sale by bureaucrats of state-owned enterprises and resources. Another major function is coordinating the payment of bribes by businesses in their efforts to expedite the regulatory process which involves 'a daunting array of license, permit, and fee requirements on normal business activity' (ibid., p. 471). The still-substantial customs procedures and high taxes also stimulate smuggling and black market trade in many consumer goods, with associated bribery of police and customs officials.

Stopping corruption

Statically, given government of a certain size, how might corruption be deterred? Most straightforwardly, by reducing the net gains to public officials from corruption. Economic actors make decisions on the basis of their values, available information and incentives. Thus, stronger incentives to become corrupt must result in more corruption. For illegal corruption, the relevant incentives are those that are delineated in the economic theory of crime: the size of expected payoffs relative to a public official's alternatives, the likelihood of being detected and punished, and the severity of the potential punishment. For legal corruption (for example, accepting legal gifts or campaign contributions in exchange for special attention, legislative votes, bureaucratic support), the incentives are similar. Punishment does not arise through the criminal justice process, but through the political process: those who lose as a consequence of the action can contribute to and support other

politicians in an effort to remove the 'offender' from office, taking away opportunities for more legal corruption in the future.

The punishment for legal corruption might seem rather different from that – jail, fines – for illegal corruption. The difference is not substantial, however, given the *expected* punishment that arises for most illegal corruption. Illegal corruption in fact goes largely unpunished, making expected punishments quite light.

Expected punishment for corruption The expected punishment for corruption is a function of two variables, (a) the joint probability of being detected and punished, and (b) the costs of the punishment to be meted out in the event of detection and punishment. It was noted above, apropos of the paucity of micro-data on corruption, that there is very little detection and punishment of corruption. That is, neither of these probabilities is very large *ex ante*, meaning that the expected costs of punishment are low.

As concerns the probability of detection, a principal question is who will monitor – and so detect – those government officials with the property rights necessary to instigate or participate in corruption. Monitoring by citizens (voters, taxpayers, interest groups) is a theoretical possibility. But practically, none of these groups has much incentive to monitor corruption. Any gains in deterring corruption go to the citizenry generally, but monitors themselves bear the costs of surveillance. Nor, second, do voter-taxpayers have ready access to the sort of information necessary to ferret out corruption. 'Except in circumstances where the problem reaches outrageous proportions, nobody monitors the progress of criminal cases to detect abuses of prosecutorial discretion; nobody raises money to support political campaigns of candidates who will eliminate police corruption; nobody watches the sentencing patterns of judges' (Neely 1982, p. 154).

The news media are also potential sources of monitoring. But they predictably will not present a major threat to most corrupt officials. Corruption exposed by others is certainly reported, but there are relatively few instances in which news personnel have actively sought out illegal activity. This is partly because newspapers and other media require daily output, and most reporters must concentrate on news that can be obtained easily and quickly. Detecting corrupt officials and proving their guilt are generally difficult and time-consuming tasks, and such efforts are likely to take place only when the potential payoff is substantial. A reporter is unlikely to invest much time and effort detecting corruption by a lower-level bureaucrat. A reporter might be willing to spend considerable time trying to demonstrate that a more important public official is corrupt because the potential payoffs are large (for example, front-page headlines, recognition by peers and citizens, and greater income opportunities), but Watergate-type reporting incidents are few and far between.

Peers can also monitor corruption, and often will have low-cost opportunities to discover it. Most governmental institutions have established self-monitoring systems, which then have actually discouraged (and in some cases even prevented) monitoring from external sources. Police departments in the United States have their internal affairs divisions, for example, and court systems have judicial review boards. Rent-extraction opportunities are everywhere in the police and judicial systems. When victimless activities like prostitution, drug use and gambling are made crimes, government officials acquire the effective property right to sell relief from arrest and fines (or even prison sentences) that would follow if they enforced the law. Small wonder that peer monitoring is commonplace among police.

But such monitoring is not likely to be very effective. There are no strong incentives to expose corruption or inefficiencies within the governmental unit. Even if internal watchdogs derive satisfaction from pursuing the public good, they face a dilemma: when they reveal corruption among their colleagues, the agency's effectiveness may well be jeopardized.

For example, the Knapp Commission on police corruption in New York attributed police officers' extreme reluctance to bring evidence against fellow officers to 'intense group loyalty', which supposedly manifested itself in a 'public-spirited' concern for the effectiveness and morale of the department. As a result, suspicion and hostility were directed at any outside interference with the department, and the 'mixture of hostility and pride created the most serious roadblock to a rational attack on police corruption: a stubborn refusal at all levels of the department to acknowledge that a serious problem exists' (Knapp 1972, pp. 6–7). Even basically honest police officers work in such a maze of rules 'that all of them must sometimes violate some of the regulations and therefore are [potentially] subject to disciplinary action. To protect themselves against such a contingency, policemen engage in a gigantic conspiracy against the outside world and cover up for each other' (Jacob 1974, pp. 10–11; see also Westley 1970; Rubenstein 1973). The police are not the only bureaucrats with strong tendencies to protect their own. Most US states have judicial review boards, for example, that involve judges in monitoring other judges, but '[s]ome critics complain … that judges cannot be counted upon to act against their own colleagues … the idea of firmly rooting out judicial corruption remains an especially sensitive one … [with] worries about the manifest danger of losing public respect' (Lacayo 1986, p. 66).

Alternatively, one can easily attribute the sort of incentives and behavior discovered by the Knapp Commission (and many other investigative commissions) to self-interest rather than public spirit. For the public official to whom power and prestige are important, revelations of corruption within the organization may lead to reductions in budget, discretionary power and prestige. In addition, uncorrupted officials may nonetheless wish to keep the corruption

option open, and so will not want to attract attention to the corruption potential of their position. The Knapp Commission (1972, p. 61) reported that 'police corruption was found to be an extensive, department wide phenomenon, indulged to some degree by a sizable majority of those on the force'.

Not surprisingly, then, the few officials who reported corruption have often been ostracized by colleagues and superiors, denied promotions, and ultimately forced to resign. When honest officials face such potential costs, it becomes even clearer that corrupt officials probably have little to fear from their peers. Therefore, 'with extremely rare exceptions, even those who themselves engage in no corrupt activities are involved in corruption in the sense that they take no steps to prevent what they know or suspect to be going on about them' (ibid., p. 3).

External monitoring by other governmental agencies is an alternative to peer monitoring. Prosecutors' offices, for instance, might seem well placed to investigate police corruption, but 'in the case of the district attorneys, there is the additional problem that they work so closely with policemen that the public tends to look upon them – and indeed they seem to look upon themselves – as allies of the police' (ibid., p. 14). The Knapp Commission found citizens had a general mistrust of the district attorneys, primarily because of these close ties. As a result of this distrust, it can be inferred that many prosecutors were also involved in the corruption (ibid., p. 5). If one is already stigmatized as corrupt anyway, the cost of actually turning to corruption is lower.

This leads to a related point: who polices the policeman's policeman? Corruption arises when someone (for example, a police officer) has the power to assign rights. As with any policing official who can 'look the other way' for a price, government officials responsible for preventing corruption by other officials have a potentially valuable right to sell. They can sell the right to practice corruption. This suggests that a corrupt official may expect the risk of punishment for corruption due to detection by another government official to be relatively low because, if detected, the corrupt official may be able to bribe the detector not to report the corruption. And not surprisingly, public officials do pay off police officers in order to practice corruption (Sherman 1978, p. 6).

Corrupt officials also may be relatively unconcerned with disclosure of their activities because external investigations are costly. In the United States, most states, counties and cities do not commit significant resources to the monitoring of public officials, relying instead on existing law enforcement agencies to monitor themselves and each other, at least until a major scandal erupts. When bureaucrats face excess demand for their services and have discretion regarding how to allocate their resources, they can easily ignore corruption, at least until it is brought to the public's attention somehow.

Whether the public is better served by use of those scarce resources in pursuit of corruption or in the provision of other services does not appear to be a question that is raised (Knapp 1972, p. 257).

Of course, there are exceptions. Operation *Greylord*, for instance, produced indictments of 30 court officials, including ten circuit judges, on charges of fixing cases, bribery, extortion, mail fraud and racketeering in Cook County, Illinois courts (Starr and Reese 1983, p. 21). But this case was unique, the culmination of an expensive three-and-a-half-year undercover investigation. Expensive efforts appear to involve a few possibly spectacular cases, such as Operation *Greylord*, perhaps in the hopes that the visibility of these actions will lead potentially corrupt officials to overestimate the risk of detection.

A more compelling explanation for these efforts is that they reflect a temporary political commitment to investigate a highly politicized scandal. Indeed, 'scandal reaction' appears to be the dominant approach to corruption control. Scandal reaction rarely addresses the true source of corruption, however. Any corruption 'clean-up' that does not address the fundamental institutional issues – the information and incentives generating the corruption – is unlikely to be successful over the long run. Without changes in the fundamental property rights and institutions, the replacements for those who are convicted, forced to resign, or defeated in an election will face the same incentives to engage in corrupt activity (Smith 1960, pp. 5–6).

Even if the probability of detection is low, corruption may be deterred if the punishment for corruption is harsh when that corruption is detected.[9] However, if officials who detect corruption in their own organizations have an incentive to suppress information and downplay its significance, then any internally generated punishment is likely to be relatively mild. Mild punishment should make the corruption appear to be less significant to those outside the organization (for example, legislators and private sector government-watch groups), thus minimizing the attention that exposure might attract. One might expect stiffer punishment when the conviction arose from detection by another organization or a private government-watch group. But anecdotal evidence suggests otherwise. Public (particularly high-ranking) officials seem to receive a short prison term and a quick parole.

> [The] Bronx County District Attorney testified before the Commission that light sentences were common in cases involving police officers. … It is clear that the risks of severe punishment for corrupt behavior are slight. A dishonest policeman knows that even if he is caught and convicted, he will probably receive a court reprimand or, at most, a fairly short jail sentence. Considering the vast sums to be made in some plainclothes squads or in narcotics enforcement the gains from corruption seem to far outweigh the risks. (Knapp 1972, pp. 252–3)

From 1970 to 1973 in New York City there was a 90 per cent turnover in the rank of captain and above, apparently due to retirement in the face of misconduct charges. But almost every criminal charge was brought against those holding the rank of lieutenant or below. The obvious implication is that punishment for police corruption is likely to be relatively light and is likely to decline as the official's rank increases.

Many American government officials would probably disagree with the preceding discussion, pointing out the stepped-up effort and success of law enforcement authorities, particularly at the federal level, in corruption detection in recent years. But committing additional resources does not guarantee their effective use. For reasons detailed above, public sector employees – who ordinarily get nothing personally as corruption detecting budgets are increased – are reluctant to report corrupt acts by their colleagues.

Expected benefits of corruption The expected benefit to a public official from corruption depends on several factors, including any preferences (or values) regarding immorality or illegality from the payoff. Potential returns to corruption will be weighed against returns to other activities that may have to be forgone if the official chooses to participate in the corrupt sale of property rights. Of course, bureaucratic officials cannot capture profits when they abstain from corruption and concentrate on enhancing efficiency in the production of whatever goods, services or rent transfers they are supposed to produce. They may be able to move to a better-paying public sector job, but few public officials receive large salaries. Furthermore, many public officials are constrained as to how and how much they can legally obtain beyond their public salaries. Thus, to the extent that public sector employment was chosen because it was an official's best alternative, a reasonably large expected payoff from corrupt activity will tempt at least some.

An obvious determinant of the payoff to corruption is the private buyer's willingness to pay for a governmental rights allocation, or the amount of wealth that could be taken away if a payoff were not made (Svensson 2003). If an official has allocative power over a number of different rights, the payoff could be large even though no single right has tremendous value. The Knapp Commission (1972, pp. 2–3) found, for instance, that 'while individual payments to uniformed men were small, mostly under $20, they were often so numerous as to add substantially to a patrolman's income'.

Furthermore, for any particular right, the greater the market distortion created by the laws being enforced, the greater the potential payoff to officials doing the enforcing. When a market is entirely outlawed, for example, as in the cases of drugs and prostitution, the potential payments to public officials for protecting a black market monopoly can be enormous. The Knapp Commission (ibid., p. 75) found evidence of payoffs to a plainclothes

police officer from gambling interests in New York to range from $400 to $1500 *per month* over 30 years ago. This is small change when compared to narcotics-related payoffs today, which can run into the hundreds of thousands of dollars.

The illicit drug market is probably the most lucrative source of police corruption that has ever existed in the United States, including the period of liquor prohibition (Ashley 1972, p. 136; Moore 1977, pp. 193–5; Rasmussen and Benson 1994, pp. 107–18). For example, Kunnes (1972, p. 43) reported, with regard to the heroin market:

> Profits are so great that corruption of law enforcement officials has become pandemic. In fact, the more officials hired for heroin suppression work, the more are bribed, or worse, become distributors themselves. Thirty federal agents within the last eighteen months alone have been indicted for being directly involved in the heroin (i.e., junk) trade.

Expansion of the cocaine market during the 1980s brought new allegations of widespread corruption. One of the judges found guilty as a consequence of Operation *Greylord* was convicted of, among other things, accepting bribes totaling $400 000 in cash and eight automobiles.

Corruption is most likely to occur when the potential payoff is high relative to the risks of detection and opportunity for legal income. Exacerbating these risks is the undercover work that is typical of drug enforcement activities. Long-term association with criminal elements facilitates corruption, in part because of a gradual erosion of the officer's value system, increased sympathy for the criminal elements, and added exposure to opportunities for illegal behavior (Girodo 1991). As Moore and Kleiman (1989, p. 2) note,

> [T]he police executive knows from bitter experience that in committing his force to attack drug trafficking and drug use, he risks corruption and abuse of authority. Informants and undercover operations – so essential to effective drug enforcement – inevitably draw police officers into close, potentially corruption relationships with the offenders they are pledged to control.

Risks of detection would appear to be modest since, as noted above, there are not strong institutional incentives to expose corruption.

Similarly, opportunities for legal earnings by police are modest. The 1990 starting salary in the largest police departments for an entry-level police officer in the United States was about $26 000; for sergeants it was about $40 000, and for chiefs of police $85 000 (Reeves 1992). In smaller departments the entry-level salaries are only slightly lower than those in the largest departments, but more experienced officers are not rewarded as well. Police chiefs in departments serving cities with a population exceeding one million receive a salary that is 3.2 times that of the entry-level officer, while chiefs in

cities with a population between 250 000 and 499 999 receive only 2.5 times the entry salary. These salaries, like those of all public servants, are modest compared to the payments that can be offered by the illicit drug industry. In 1990, the markup on one kilogram of cocaine at the wholesale level in the highest cost markets was $135 000 – enough to pay the annual salary of the chief of police, a sergeant and offer $10 000 to a rookie officer (DEA 1991). Drug traffickers were reported to offer $100 000 to DEA officers 'for openers', and one high US Border Patrol official reported an offer of $5 million. High returns can be achieved with little effort. In 1986 a federal prosecutor was charged with receiving payments of $210 000 and a boat in exchange for tipping off a drug smuggler to the evidence-gathering activities of US DEA officers (Press and Starr 1986, p. 68). Several examples of police officers seizing drugs and then selling them have been documented.[10]

Relative payoffs can be quite high in other areas as well. Recent evidence from Uganda (Svensson 2003) reveals that payoffs from firms paying bribes average $8300 per year (in a country whose per capita GDP is about $1000), some 8 per cent of a firm's total costs. The Knapp Commission (1972, p. 2) found that investigating detectives' 'shake-downs of individual targets of opportunity' frequently 'come to several thousand dollars', for instance. Cho and Kim (2001) report large payoffs for regulators in Korea via employment in regulated firms after regulatory bureaucrats retire.

If the power to influence a rights assignment is widely dispersed and difficult to coordinate, however, the payoff to any one official is likely to be relatively small. Organized crime may have to bribe several police officers, for instance, to ensure the relatively unmolested operation of their underground markets in drugs and prostitution, but this means that the payoff to any one police officer will be relatively small and less acceptable. Similarly, if a buyer of arbitrarily reallocated rights has several alternative sources (competitive corruption, if you will), then the return to any one corrupt seller is likely to be small.

Conclusion

Any persuasive economic analysis of corruption – positive or normative – must begin with a model of government. 'Without understanding why the state exists, it is difficult to assess why corruption arises, what its consequences are, and whether and how it should be prevented' (Acemoglu and Verdier 1998, p. 1381). Those who cite the negative correlation between corruption and economic growth or investment (for example, Rose-Ackerman 1999, p. 2) to support their contention that all corruption is bad fail to recognize that both high levels of corruption and low levels of economic growth/investment arise because government officials control the allocation of many property rights.

Indeed, there is a more fundamental normative question: should government officials have the discretionary power to allocate property rights? Fully addressing the various issues that bear on this question is beyond the scope of this chapter, but the fact that corruption is an inevitable consequence of such power has implications for normative analysis. Indeed, those who see corruption as something that is always bad should be at least normatively biased in favor of limited government powers, since that is likely to be the only effective way to eliminate corruption. Furthermore, it is suggested here, given the discretionary power that government officials have, corruption can be either good or bad depending on the circumstances. Given such power, however, a property rights perspective provides a clear way to distinguish economically between the determinants of good and bad corruption.

Notes

* Research assistance from Jim McMasters and Suzette Won is acknowledged with gratitude.

1. Jimmy Hughes, 'Steal Away' (Screen Gems – EMI Music, Inc., BMI). Hughes wrote and performed the song, which reached number 17 on the American Billboard charts in July 1964.
2. Svensson (2003) is remarkable for its use of firm-level data on bribery collected privately, concerning corruption that was otherwise not reported or prosecuted.
3. Laws 'have proliferated so rapidly as to suggest (even to lawyers) that American society is choking from legal pollution' (Auerbach 1983, p. 9). Similarly, Leoni (1961, p. 145) noted, in discussing the written codes that provided the basis of several legal systems in Western Europe, that 'gradually, the original closed system of the codes became surrounded and overburdened with an enormous system of other [legislated] rules, the accumulation of which is one of the most striking features of present-day European legal systems'.
4. As Friedman (1973, pp. 180–81) writes, 'Imagine buying cars the way we buy governments. Ten thousand people would get together and agree to vote, each for the car he preferred. Whichever car won each of the ten thousand would have to buy it. It would not pay any of us to make any serious effort to find out which car was best; whatever I decide, my car is being picked for me by the other members of the group. Under such institutions the quality of cars would quickly decline. That is how I must buy products on the political marketplace. I not only cannot compare alternative products, it would not be worth my while to do so even if I could. This may have something to do with the quality of the goods sold on that market. *Caveat emptor*'.
5. Bureaucrats can also shirk and cut quality because of the lack of effective monitoring. Sherman's (1983, p. 151) extensive review of research on police performance suggests that about half of a typical patrol officer's time is spent simply waiting for something to happen, for example. While police officials claim that this time is spent in preventive patrolling, systematic observation indicates that such time is largely occupied with conversations with other officers, personal errands, and sitting in parked cars on side streets.
6. Public court backlogs are tremendous in some states, for instance. It can take four or more years to get an auto accident case heard in California. Delay is costly for the person who cannot collect damages to pay for ongoing expenses, but anxiety and frustration costs can also be significant. Police files are also full of reported crimes waiting, many forever, to be solved, costing victims who want justice dearly as they must continue to pressure police to act. Similarly, public prisons are so crowded that most prisoners never serve their full sentence while convicted criminals queue up in local jails waiting for a place in state prisons, raising costs to local taxpayers, to past victims who do not see satisfactory

punishment and to future victims because of the reduced incapacitation and deterrent effects.

7. Common-pool analysis is usually applied to natural resources. However, applying it to publicly provided services, including those provided corruptly, can provide considerable insights, particularly with regard to courts and the criminal justice system (for example, Shoup 1964; Neely 1982; Barnett 1986; Benson 1988b, 1990, 1994; Benson and Wollan 1989; Benson and Rasmussen 1991; Rasmussen and Benson 1994; Ekelund and Dorton 2003).
8. Organized crime arose in Eastern Europe during the communist era, often to supply consumer goods and services to government officials, who were among the few with sufficient financial resources to buy them. The fall of the totalitarian regimes did not end organized crime, however, or government ties to it, because of the strong controls that the government still maintains over economic activities. Indeed, the power of organized crime in Russia (and elsewhere, as argued in Benson 1988a) still emanates from government, and generally to the substantial benefit of those in government.
9. The impact of punishment is difficult to assess, however, since severity is a subjective concept. An official who obtains satisfaction from a prestigious position may view the embarrassment of public exposure for corruption, and the loss of a job, as severe punishment; another with attractive outside alternatives might view the exposure as an inconvenience.
10. For instance, the former sheriff of Nassau County, Florida, was convicted in July 1993 of distributing drugs from his department's property room. He reportedly pretended to destroy marijuana and cocaine but sold it instead. In addition, he actually requested 18 pounds of cocaine and 100 pounds of marijuana from other counties for use in 'undercover' investigations by his partner in crime – one of his deputies – and then reported that the operations fell through and the drugs were destroyed, when in fact they were apparently sold. The sheriff supposedly earned between $175 000 and $225 000 from these drug sales (Ward 1993, p. 2C).

References

Abbot, K.W. and D. Snidal (2002), 'Values and interests: international legalization in the fight against corruption', *Journal of Legal Studies*, **31** (1), 141–78.

Acemoglu, D. and Th. Verdier (1998), 'Property rights, corruption and the allocation of talent: a general equilibrium approach', *Economic Journal*, **108**, 1381–5.

Acemoglu, D. and Th. Verdier (2000), 'The choice between market failures and corruption', *American Economic Review*, **90** (1), 194–211.

Ades, A. and R. Di Tella (1997a), 'The new economics of corruption: a survey and some new results', *Political Studies*, **45**, 496–515.

Ades, A. and R. Di Tella (1997b), 'National champions and corruption: some unpleasant interventionist arithmetic', *Economic Journal*, **107**, 1023–42.

Alam, M. Shahid (1990), 'Some economic costs of corruption in LDCs', *Journal of Development Studies*, **27**, 89–97.

Ashley, Richard (1972), *Heroin: The Myths and Facts*, New York: St. Martin's Press.

Auerbach, J.S. (1983), Justice without Law?, New York, Oxford University Press.

Bardhan, P. (1997), 'Corruption and development: a review of issues', *Journal of Economic Literature*, **35** (3), 1320–46.

Barnett, R.E. (1986), 'Pursuing justice in a free society, Part Two: Crime prevention and legal order', *Criminal Justice Ethics*, **5**, 30–53.

Basu, K., S. Bhattacharya and A. Mishra (1992), 'Notes on bribery and the control of corruption', *Journal of Public Economics*, **48**, 349–59.

Becker, G.S. (1968), 'Crime and punishment: an economic approach', *Journal of Political Economy*, **76**, 169–217.

Benson, B.L. (1981), 'A note on corruption of public officials: the black market for property rights', *Journal of Libertarian Studies*, **5** (3), 305–11.

Benson, B.L. (1984), 'Rent seeking from a property rights perspective', *Southern Economic Journal*, **51** (1), 388–400.
Benson, B.L. (1985), 'Understanding bureaucratic behavior: implications from the public choice literature', *Journal of Public Finance and Public Choice*, **8**, (2–3), 89–117.
Benson, B.L. (1988a), 'An institutional explanation for corruption of criminal justice officials', *Cato Journal*, **8** (1), 139–63.
Benson, B.L. (1988b), 'Corruption in law enforcement: one consequence of "The tragedy of the commons" arising with public allocation processes', *International Review of Law and Economics*, **8**, 73–84.
Benson, Bruce L. (1990), *The Enterprise of Law: Justice Without the State*, San Francisco, CA: Pacific Research Institute for Public Policy.
Benson, B.L. (1994), 'Are public goods really common pools: considerations of the evolution of policing and highways in England', *Economic Inquiry*, **32** (2), 249–71.
Benson, B.L. (2002), 'Regulatory disequilibrium and inefficiency: the case of interstate trucking', *Review of Austrian Economics*, **15** (2–3), 229–55.
Benson, B.L. and J. Baden (1985), 'The political economy of government corruption: the logic of underground government', *Journal of Legal Studies*, **14** (2), 391–410.
Benson, B.L. and D.W. Rasmussen (1991), 'The relationship between illicit drug enforcement policy and property crimes', *Contemporary Policy Issues*, **9** (4), 106–15.
Benson, B.L. and L.A. Wollan, Jr. (1989), 'Prison overcrowding and judicial incentives', *Madison Paper Series*, **3**, 1–21.
Bicchieri, C. and J. Duffy (1997), 'Corruption cycles', in Heywood (ed.), pp. 61–79.
Bowles, R. (2000), 'Corruption', in Boudewijn Bouckaert and Gerrit De Geest (eds), *Encyclopedia of Law and Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 460–91.
Breton, Albert and Ronald Wintrobe (1982), *The Logic of Bureaucratic Control*, Cambridge: Cambridge University Press.
Buchanan, James M. and Gordon Tullock (1962), *The Calculus of Consent*, Ann Arbor, MI: University of Michigan Press.
Cheung, S.N.S. (1996) 'A simplistic general equilibrium theory of corruption', *Contemporary Economic Policy*, **14** (3), 1–5.
Cho, J. and I. Kim (2001), 'Jobs in the bureaucratic afterlife: a corruption-facilitating mechanism associated with law enforcement', *Southern Economic Journal*, **68** (2), 330–48.
de Soto, Hernando (1989), *The Other Path: The Invisible Revolution in the Third World*, New York: Harper & Row.
de Soto, Hernando (2000), *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else*, New York: Basic Books.
Dempsey, G. and A. Lukas (1998), 'Mafia capitalism or red legacy?', *Freeman*, **48**, 469–72.
Demsetz, H. (1967), 'Toward a theory of property rights', *American Economic Review*, **57**, 347–59.
Drug Enforcement Administration (DEA) (1991), *Illegal Drug Price/Purity Report, United States 1981–1990*, Washington, DC: United States Department of Justice, DEA.
Ekelund, R.B. and C. Dorton (2003), 'Criminal justice institutions as a common pool: the nineteenth century analysis of Edusin Chadwick', *Journal of Economic Behavior and Organization*, **50**, 271–94.
Eggertsson, Thráinn (1990), *Economic Behavior and Institutions*, Cambridge: Cambridge University Press.
Fiorentini, Gianluca and Stefano Zamagni (eds) (1999), *The Economics of Corruption and Illegal Markets*, 3 vols, International Library of Critical Writings in Economics, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Friedman, David (1973), *The Machinery of Freedom: Guide to Radical Capitalism*, New York: Harper & Row.
Girodo, M. (1991), 'Drug corruption of undercover agents: measuring the risk', *Behavioral Sciences and the Law*, **9**, 361–70.
Goel, R.K. and M.A. Nelson (1998), 'Corruption and government size: a disaggregated analysis', *Public Choice*, **97**, 107–20.

Heidenheimer, A.J. (1996), 'The topography of corruption: explorations in comparative perspective', *International Social Science Journal*, **48** (3), 337–47.
Heywood, Paul (ed.) (1997), *Political Corruption*, Oxford, UK: Blackwell.
Huntington, Samuel (1968), *Political Order in Changing Societies*, New Haven, CT: Yale University Press.
Jacob, Herbert (ed.) (1974), *The Potential for Reform of Criminal Justice; Sage Criminal Justice System Annuals*, vol. 3, Beverly Hills, CA: Sage.
Kaufmann, D. (1998), 'Corruption in transition economies', in Peter Newman (ed.), *The New Palgrave Dictionary of Economics and the Law*, vol. I, London: Macmillan Reference and New York: Stockton Press, pp. 522–30.
Klitgaard, R. (1998), 'International cooperation against corruption', *Finance and Development*, **35**, 3–6.
Knapp, Whitman (1972), *The Knapp Commission Report on Police Corruption*, New York: George Braziller.
Krueger, A.O. (1974), 'The political economy of a rent-seeking society', *American Economic Review*, **64**, 291–303.
Kunnes, Richard (1972), *The American Heroin Empire: Power, Profits, and Politics*, New York: Dodd, Mead.
Lacayo, R. (1986), 'Passing judgement on the judges: a spate of legal trouble for the judiciary', *Time*, 20 January, 66.
Leff, N.H. (1964), 'Economic development through bureaucratic corruption', *American Behavioral Scientist*, **8** (3), 8–14.
Leoni, Bruno (1961), *Freedom and the Law*, Los Angeles, CA: Nash.
Mauro, P. (1995), 'Corruption and growth', *Quarterly Journal of Economics*, **110** (3), 681–712.
Mauro, P. (1998), 'Corruption: causes, consequences, and agenda for further research', *Finance and Development*, **35**, 11–14.
McChesney, F.S. (1987), 'Rent extraction and rent creation in the economic theory of regulation', *Journal of Legal Studies*, **16** (1), 101–18.
McChesney, Fred S. (1997), *Money for Nothing: Politicians, Rent Extraction, and Political Extortion*, Cambridge, MA: Harvard University Press.
Mookherjee, D. and I.P.L. Png (1995), 'Corruptible law enforcers: how should they be compensated?', *Economic Journal*, **105** (1), 145–59.
Moore, Mark H. (1977), *Buy and Bust: The Effective Regulation of an Illicit Market in Heroin*, Lexington, MA: Lexington Books.
Moore, M.H. and M.A.R. Kleiman (1989), 'The police and drugs', *Perspectives on Policing*, No. 11, Washington, DC: National Institute of Justice.
Neely, Richard (1982), *Why Courts Don't Work*, New York: McGraw-Hill.
Olson, Mancur (1965), *The Logic of Collective Action*, Cambridge, MA: Harvard University Press.
Press, A. and M. Starr (1986), 'The friends of David T: a prosecutor on trial', *Newsweek*, 10 February, 68.
Rasmussen, David W. and Bruce L. Benson (1994), *The Economic Anatomy of a Drug War: Criminal Justice in the Commons*, Lanham, MD: Rowman & Littlefield.
Reeves, B.A. (1992), 'State and local police departments, 1990', *Bureau of Justice Statistics Bulletin*, February, 1–14.
Rose-Ackerman, Susan (1999), *Corruption and Government: Causes, Consequences and Reform*, Cambridge, UK and New York: Cambridge University Press.
Rubenstein, Jonathan (1973), *City Police*, New York: Farrar, Straus, Giroux.
Sherman, Lawrence W. (1978), *Controlling Police Corruption*, Washington, DC: National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration.
Sherman, Lawrence W. (1983), 'Patrol strategies for police', in James Q. Wilson (ed.), *Crime and Public Policy*, San Francisco: Institute for Contemporary Studies.
Shoup, C. (1964), 'Standards for distributing a free government service: crime prevention', *Public Finance*, **19**, 383–92.
Simons, W.B. (1996), 'Corporate law in the private sector: the role of state control', in G. Ginsburgs, D.D. Barry and W.B. Simons (eds), *The Revival of Private Law in Central and*

Eastern Europe: Essays in Honor of F.J.M. Feldbrugge, The Hague: Martinus Nijhoff, pp. 249–80.
Smith, Bruce (ed.) (1960), *Police in the United States*, 2nd rev. edn, New York: Harper & Row.
Starr, Mark and Michael Reese (1983), 'Stinging the Chicago Courts', *Newsweek*, **102**, 22 August, 21.
Stratmann, Th. (1991), 'What do campaign contributions buy? Deciphering causal effects of money and votes', *Southern Economic Journal*, **57**, 606–20.
Stratmann, Th. (1992), 'Are contributors rational? Untangling strategies of political action committees', *Journal of Political Economy*, **100** (2), 647–64.
Stratmann, Th. (1996), 'How reelection constituencies matter: evidence from political action committees' contributions and congressional voting', *Journal of Law and Economics*, **39** (2), 603–36.
Stratmann, Th. (1998), 'The market for congressional votes: is timing of contributions everything?', *Journal of Law and Economics*, **41** (1), 85–114.
Svensson, Jakob (2003), 'Who must pay bribes and how much? Evidence from a cross section of firms', *Quarterly Journal of Economics*, **118**, 202–30.
Tanzi, V. and H. Davoodi (1998), 'Roads to nowhere: how corruption in public investments hurts growth', *IMF Economic Issues*, No. 12.
Thornton, Mark (1991), *The Economics of Prohibition*, Salt Lake City, UT: University of Utah Press.
Tullock, G. (1967), 'The welfare costs of tariffs, monopolies and theft', *Western Economic Journal*, **5**, 224–32.
Tullock, Gordon (1993), *Rent Seeking*, Aldershot, UK and Brookfield, US: Edward Elgar.
Ward, Ron (1993), 'Ex-Sheriff convicted in drug case; faces sentence of life in prison', *Tallahassee Democrat*, 16 July, 2c.
Westley, W.A. (1970), *Violence and the Police*, Cambridge, MA: MIT Press.

16 Insider trading, takeovers and property rights

Pierre Garello

Why should we deal with these two issues and what relates them?

The first common point is surely that both issues, insider trading and takeover, have attracted much attention, not only from the media but also from regulators and legislators, and consequently, from scholars. Today, insider trading and takeovers are among the most regulated property rights transactions. Moreover, those regulations have been and still are the object of many controversies.

Another common point is that both issues are related to corporate governance. Insider trading, at least initially, describes a type of behaviour on the part of top employees that was judged to be detrimental to the owners of the company, that is, the shareholders; while the goal of a takeover is most of the time to change the management of a corporation, or at least to modify its strategy. Hence at the heart of both issues we find the relationship uniting shareholders and managers, even though, as will be recalled below, the general functioning of the market is also a matter of concern for the regulator.

A third common point relates to the problem of jurisdiction. In both cases the search for a solution to the problem could either be left to the parties directly involved in the contractual relationship (the contract between managers and owners, the contract between co-owners, or the contract between companies and trading centres), or entrusted to the local state; a third possibility, is that it could be left to the federal level (for example in the United States) or at the level of the Union in Europe. The tendency, again in both cases, has been towards 'centralization', 'forced harmonization' and 'federalization', rather than competition. One of the aims of the present study is to evaluate that tendency: does it go in the right direction? To answer this question we shall need to specify what a good direction would be, which in turn raises the preliminary question of what is meant by a proper functioning of the market.

Finally, the two topics are naturally linked by the fact that most insider trading violations appear to be related to takeovers.[1]

To address these issues properly and evaluate related laws and regulations requires first a reflection on the nature of the title of ownership bought by a shareholder, and in particular on the status of limited liability. The desire to regulate or prohibit insider trading and takeovers comes largely from a lack of consideration for the reasons that initially led the parties to choose a

regime of co-ownership and limited liability. Following this general reflection on limited liability, we shall examine respectively the economics and law of insider trading and then of takeovers. In a final section, we shall leave aside the more theoretical questions to ask ourselves who has a direct interest in such regulations? This public choice perspective might give us a better prediction on the future of those rules and regulations that govern insider trading and takeovers.

Limited liability: its origins and consequences

What types of property rights are traded by an insider or during a takeover? Those are titles that make their bearers co-owners of a company, most of the time with limited liability. It is interesting to reflect on the reasons why such a specific form of ownership developed in the first place.[2] At the beginning was the need to raise significant amounts of capital. Due to the size of the investment, the enterprise was too risky for a single investor. Hence the desire came to develop some legal concepts of partnership, and indeed many new legal arrangements have been invented, from the twelfth century to the present. But each 'solution' had its own drawbacks. In particular, as more partners were called into a project, and as the project increased in size, liability became a preoccupation. Each partner had less control on the way the business was run as well as on the way his/her partners were handling their own business.

Limited liability can be seen as a possible response to the control problem that resulted from the increased number of partners.[3] It is easier to convince investors – at least some of them – to invest capital in a company if they know their liability is limited. Hence, limited liability was a compensation offered for joining a business in which one has little control. Nonetheless, this contractual solution came, as always, at some cost: to third parties and to the co-owners themselves.

Third parties include clients of the company (consumers, suppliers), as well as lenders to the company. Following the introduction of limited liability, third parties have to increase their control of the company precisely because co-owners, being made less liable, have less incentive to exercise that control themselves.[4] Lenders, therefore, will most probably ask for higher rates while clients will ask for lower prices, both situations leading, everything else equal, to lower returns for shareholders. Hence, limited liability shifted risk, and it is customary for those who bear more risk to demand a premium that will be paid by those who have reduced their exposure to risk, namely, the shareholders.

Being imperfect, the limited liability option is not always the most interesting form of co-ownership. An alternative could be to stick to unlimited liability. In that case, co-owners will have to diversify risk themselves. They

will most probably choose to rely on 'internal diversification' – the same company diversifying its activities – rather than 'external diversification' – individuals buying shares in a variety of companies which are each rather undiversified. Furthermore, in an unlimited liability company, co-owners will not only have greater incentives to check on management's behaviour and on the degree of risk involved in the project they undertake, but they will also have to control the behaviour of the other co-owners, their liabilities being tied.

Which option – limited or unlimited liability – is the best probably depends on time and place, on whether external or internal diversification is more appropriate, on the degree of risk aversion of market participants, and, last but not least, on the availability and cost of insurance. The point needing to be stressed for the purpose of our inquiry on insider trading and takeovers is that by choosing limited liability, as many companies did, the choice was made, so to speak, to 'get around' the control problem instead of trying to deal with it in a more direct way. One can hardly opt for limited liability while simultaneously keeping a high degree of control. There may be many ways of diversifying risk, but there is no way to eradicate it, at least as long as we are participating in a dynamic, changing economy.

Finally, it should be recalled that, besides limited liability, another complementary strategy has been adopted to raise more funds: to reduce substantially the cost of exit. Nowadays, in many instances it is possible to sell shares by means of a simple phone call, no authorization being required from other shareholders or managers. Clearly, when such a strategy is used to raise capital a worsening of the control problem is to be expected.

However, although no institutional arrangement is perfect, we should not embark on the slippery slope of resignation. Property rights arrangements can and do evolve as will be shown below; and many ways exist and are still being invented in order to improve the simplest form of limited liability. It remains nonetheless evident that limited liability and the low cost of exit constitute a handicap for those who are seeking tight control over choices made by the company's management, including insider trading and takeovers.

Insider trading

Insider trading occurs when an insider is trading in securities while in possession of material non-public information. In the United States, Section 16 (a) of the Securities Exchange Act of 1934 started by defining an insider as an officer, director or anyone who holds more than 10 per cent of equity shares. That definition, however, was judged to be unsatisfactory because someone in possession of a valuable piece of information, without holding one of the positions specified by the law, could trade without having to worry about insider trading regulation. Later, the definition was therefore broadened, so

much so that, as we shall see, the regulation on insider trading has been applied to 'outsiders' who were trading on the basis of information which was not even coming directly from inside the company. What a misnomer!

Leaving aside the definition of an insider, the general idea is that the private use of inside information would be detrimental to shareholders' interests and undermine market morality. As a consequence, in order to protect actual shareholders and attract new ones, either information should be made public or its use should be illegal. Before developing and examining this argument, one can, in the light of our previous discussion, question its robustness. Indeed, if the intention with limited liability was to avoid rather than solve the control problem, then shareholders cannot reasonably expect that their trading partners will never be better informed than themselves. Hence, the promise formulated by legislators and regulators to ensure relatively equal access to information to all traders, appears, at first sight, unreasonable. Below, the logic of this approach will be further studied before we examine the history of legal prohibition.

To start with, there is no evidence that insider trading drives investors away from the market. As a matter of fact, while more insider trading cases were brought to the attention of the general public in the 1980s and 1990s, securities market capitalization kept growing.[5] An explanation for this lack of reaction to insider trading could simply be that there is little to react about: shareholders' interest would hardly be affected by such behaviour. So, let us look more carefully at the likely consequences of insider trading for shareholders.

Shareholders could be injured because part of the gains associated with 'good news' for their company will go to the insider rather than to them. This could be especially true when they *sell* their shares to insiders. More precisely, in order to grasp their profit opportunities, insiders would have to trade with some shareholders of their company, and those shareholders will therefore reap little advantage from the good news.

Without denying this fact, it must be emphasized that the damage thereby caused to shareholders is neither so great nor so immoral for the following reasons. First, insiders will trade only on a limited volume of shares due to their wealth constraint and to their fear of being unmasked, thus, only a small portion of shareholders will lose the opportunity to gain from the good news. In the meantime, the other shareholders, those who do not sell to the insider, will benefit from insider trading because, assuming that insider trading has an effect on the price of the share, it will precipitate a price increase. Second, it must be recalled that those shareholders who sell to the insider do it voluntarily: they were looking for a buyer when the insider offered to trade. This means that either they thought no good news was likely to come from that company – in which case they should blame themselves for holding erro-

neous expectations – or they had some liquidity problem that the insider helped to solve. Without the insider they would have had to wait or to sell at an even lower price.

The damage to shareholders could well be limited for another, indirect, reason. As pointed out by Manne (1966), the possibility of trading on inside information gives further incentives for managers to 'produce' good news. Gains realized by insiders would hence compensate them for their effort; the compensation being positively correlated with the increase they generate in the assets' value, and therefore with the increase in shareholders' wealth. This argument is, however, weakened by two considerations. On the one hand, the insider is not necessarily the producer of the good news; he/she is not necessarily responsible for the value increase. On the other, the material non-public information on the basis of which the insider trades is not always goods news for the company; it can as well be bad news. When looking at the two sides of insider trading, it appears then that, as far as incentives are concerned, allowing insider trading will tend to push managers to undertake riskier projects since it is possible for them to grasp the potential profits without having to bear the potential losses. Now, whether or not it is desirable to have more risk and potentially higher returns is a matter of subjective preference, and surely shareholders' preferences will in general differ: some may welcome a riskier strategy, while others may wish to avoid it.

Bad news for the company, as said above, will lead the insider in possession of that piece of information to sell his/her shares either to someone who already owns some, or to some 'outside' investors. The first case is similar to the one we just dealt with. The second case is different because here the insider is not, prior to trade, related to the outsider and, consequently, there can be no breach of fiduciary duty through trade. The case for prohibition is therefore weakened. Nonetheless, even though no fiduciary duty can be established it has been argued that the market would be more efficient and attract more participants if insider trading were prohibited. This line of argument, however, reveals, in our opinion, a misunderstanding of the functioning of the market. Markets are not efficient to the extent that each trader is perfectly informed (Fama's approach); they are efficient *because* they allow us to trade on the basis of our personal and incomplete knowledge, thereby giving us an incentive to develop new knowledge (see Hayek 1945). Indeed, the strength of the market system is that it promotes the expansion of knowledge through a greater division of knowledge. The battle for equal access to information for all traders is therefore not only a lost battle, but a battle that is more likely to reduce efficiency.

While in agreement with the above analysis of the market system, some may add that the market system is nothing but an institutional arrangement based on property rights and contracts, and that it will perform its task even

better if information is protected by property rights. Hence, insider trading regulation will increase efficiency, just as patents, copyrights and trademarks do.

To see insider trading regulation as defining *de facto* property rights on information is, however, problematic. The problem arises when one must clarify who owns the rights. If they belong to shareholders then the insider who owns some shares, also owns the rights. Therefore it is not enough to point out that shareholders have property rights on information to ban insider trading; one must also add that those who own the rights cannot trade on the basis of these rights. But then the economic advantage commonly attached to the definition of property rights vanishes. The most obvious solution is therefore to allocate the property rights to the company. But here again the economic argument is weak. Indeed, contrary to the case of patents and copyrights, the insider's use of information does not necessarily harm the company. In some cases, such as takeovers, the company can even benefit from insider trading which increases the chance of success of the takeover. Hence, there is no clear economic argument in favour of an assignment of property rights on non-public information.

To end our theoretical inquiry let us add that, even if one agrees that insider trading has to be banned, the question still remains: who will be in the best position to do it, and at what cost? It is sometimes argued that economies of scale are realized when the control is entrusted to a general administration such as the Securities Exchange Commission in the United States, but even those studies that examine the effectiveness of insider trading regulation do not take into account the administrative cost of such a regulation.[6] The economies-of-scale argument is therefore far from being conclusive. Also, could insider trading be dealt with at the level of the firm? Would not a contractual solution to that agency problem be as efficient – from a cost–benefit point of view – as an administrative control? The company desirous of signalling to its shareholders that something is being done in the company in order to avoid insider trading could then hire a private agency that would do the work currently performed by the administration. Insider trading cases would then be treated just like any other breach of contract.

Turning now to the legal aspect of the problem, our first observation is that law and regulation have been using various arguments since insider trading became an issue (for references, see Bainbridge 2001). Prior to the 1934 Securities and Exchange Act, insider trading cases were in the competence of state laws and there was little prohibition. Of course, fraudulent concealment or misrepresentation were a matter of liability, but, in the absence of such behaviour, the rule was that the officers of the company had no duty to disclose all the information they might have about the company before trading in their company's shares. In 1903, the Georgia Supreme Court reversed

this rule by establishing a duty to disclose for officers; but a 1909 US Supreme Court's decision recognized that such a duty could exist only under special circumstances, and furthermore, only for the purchase of shares, since there could be no fiduciary duty with respect to non-shareholders to whom you sell shares.

The year 1934, however, marked a complete change in the way insider trading was to be legally handled in the future. The Securities and Exchange Commission (SEC) was created 'to protect investors and increase public confidence in the reliability of the securities markets'. To do so, it was believed necessary to go beyond the mere condemnation of fraudulent behaviour and to impose full disclosure of information by insiders wishing to trade. The 1934 Act, however, did not target insider trading as one of the behaviours to be banned and, as a matter of fact, insider trading was not explicitly mentioned in the law. The closest it comes to mentioning it was probably with Rule 10b, which states that it is unlawful 'to use any manipulation or deceptive device or contrivance or contravention of such rules and regulation as the commission may prescribe as necessary or appropriate in the public interest or for the protection of investors'.[7] Even though that rule was completed in 1942 (Rule 10b-5), the prohibition of insider trading remained at that time limited to face-to-face transactions.

A new turn occurred in 1961 when the SEC ruled that insider trading on an impersonal stock exchange was violating Rule 10b-5.[8] Shortly after, the Second Circuit Court of Appeal confirmed the SEC decision and developed the 'disclose or abstain' rule, according to which insiders must either disclose all relevant information they possess or abstain from trading in the shares of their company. The decision was founded on the necessity, according to the judges, of ensuring that all traders have 'relatively equal access to material information'.

But this was far from ending the legal controversy. Indeed, in the early 1980s, the Supreme Court temporarily put a hold on a trend of ever-increasing prohibition. In the *Chiarella* case it reversed a decision that condemned someone who traded on the basis of privileged information but was not an employee, while in the *Dirk* case, Dirk, a securities analyst who had tipped off some clients, was also declared innocent.[9] The SEC, however, did not remain silent and its reply came soon after with the addition of a new rule, Rule 14e-3. Under this rule it is forbidden to release information about a tender offer to someone who is likely to trade on the basis of that information for his/her own interest. Besides Rule 14e-3, which applies only to tender offer transactions, the doctrine of 'misappropriation' was also developed. According to that doctrine, an outsider, although not violating any fiduciary duty to the person with whom he or she trades, can nonetheless be convicted of infringing a fiduciary duty to the source of the information. Then in 1984,

Congress passed the Insider Trading Sanction Act, followed in 1988 by the Insider Trading and Securities Fraud Enforcement Act, both aimed at an enhanced compliance with insider trading regulation in the United States.[10]

As one can see, most of the regulation of the last 50 years was initiated at the federal level. State corporate law, however, is not totally silent on the question. Indeed, one can rely on agency law and the fiduciary duty it acknowledges between the principal and the agent. Hence, to give an illustration, Section 395 states that the agent may not use for personal gain any information given to him/her by the principal or acquired by him/her during the course of or on account of his/her agency. But agency law also requires that the principal be injured by the behaviour of the agent, which, as recalled above, is often not the case.

Having surveyed both the law and the economics of insider trading, we are left with a feeling of disproportion: the fierce attacks on insider trading launched by the administration, with the support of most legislators, appears out of proportion to the problem potentially raised by insider trading. Because, without denying that insiders may harm some shareholders, insider trading can just as easily be beneficial to many other shareholders and to the company as well, so a cost–benefit analysis can hardly be conclusive. Now, if instead of a cost–benefit analysis one wishes to justify the prohibition with a fiduciary duty towards one's employer – a justification which, as we argued, is weak – this will limit the prohibition to cases in which insiders have themselves been purchasing shares, thus injuring the company, or some of the shareholders. Finally, to argue that the insider is 'stealing' the information from the company, thereby violating property rights is not satisfactory, because the economic incentives to define and enforce such rights are doubtful.

Takeovers

To take over a company is one way to gain control over it and often, to dismiss the incumbent management.[11] However, many takeovers are negotiated with the incumbent management; and when they are not, they can nonetheless receive its approval (friendly takeover), so that takeovers met with fierce resistance (hostile offer) are rather the exception. For reasons that will be developed below – related mainly to the capital structure of publicly traded corporations – it is mainly in the United States and the UK that takeovers have attracted a lot of public attention and raised many questions such as: what motivates those changes in control, do they promote economic development, and, consequently, should they be regulated?

Takeovers are heavily regulated and if this regulation is nowadays closely tied to insider trading regulation, such has not always been the case. Historically, the first regulation of takeovers was conceived mainly as a safeguard for consumers and small businesses rather than for shareholders. In other

words, the problem was seen not with the transaction *per se* – one party to the transaction requiring protection – but with the distribution of power among market participants that could result from those transactions; a distribution of power that was judged to be detrimental to 'social welfare'. We shall start with a critical examination of this line of argument against takeover before dealing with the more recent and more corporate governance-oriented aspect of takeover regulation.

At the end of the nineteenth century, important innovations in corporate law took place (supported in particular by the states of Delaware and New Jersey), giving a good illustration of the process of institutional innovation, and more particularly of innovation in property rights arrangements. The Sherman Act, a federal law, was then passed essentially out of fear that some of those innovations would impede on interstate businesses.[12] It was thought that large cartels could destroy competition. At that time, most economists saw essentially positive aspects to mergers, which were judged to be the source of economies of scale and therefore to enhance economic efficiency. Indeed, to confirm that this first piece of legislation was not directed against mergers, it is enough to recall that between 1898 and 1902 there was a great wave of mergers.

After the election of Theodore Roosevelt, hostility towards mergers grew. However, this did not last and, in the 1920s, as enforcement of antitrust law was somehow scaled back, a new wave of mergers and acquisitions took place, giving rise to the modern corporation as we know it, with a great number of shareholders and a new allocation of power between managers and shareholders. It is this evolution that motivated A.A. Berle and G.C. Means to write their famous book *The Modern Corporation and Private Property* (1932) in which they criticized such a development.

As in the case of insider trading, the regulation framework changed and became largely a matter of federal law in 1934 when the Securities Exchange Act was passed. Indeed, 1934 marked the beginning of a long period of hostility towards takeovers which came to be seen as means to obtain monopoly power; monopoly power being in turn perceived as incompatible with competition and therefore suboptimal in the neoclassical–Paretian sense of the word. The Williams Act passed in 1968 completed the work by imposing even stricter disclosure requirements and restricting the terms of a tender offer.[13] Hence, by the end of the 1960s, market authorities were typically forbidding a merger involving market shares of as little as 5 per cent of the 'relevant market'.[14] Not surprisingly, however, takeovers did not disappear; rather, they changed track: it is probably due to those regulations that conglomerates, which sometimes had little economic sense, were formed.[15]

If in the 1980s most economists still considered the degree of concentration of a 'given' market as a threat to economic efficiency, the judgement,

however, was now more cautious.[16] Hence, any takeover – and more generally any business practice – likely to lead to a 'substantial lessening of competition' had to be checked out and eventually forbidden. The era of a priori control and of the 'rule of reason' hence entered the stage. According to that rule, prior to any takeover, experts mandated by the market authorities must check whether or not the transaction would be to the benefit of the whole community (see Kovacic and Shapiro 2000).

From a theoretical point of view, however, such an approach raises new problems. To ask experts to weigh social costs (essentially, monopoly power) and social benefits (essentially economies of scale, more rapid spreading of new technology, or business synergies) can hardly be reconciled with a dynamic understanding of the market. Indeed, the neoclassical paradigm on which that regulation is in great part founded could soon start to stumble under the attack of competing views that saw the market as a discovery process whose outcome is therefore unpredictable – even to the best expert – and monopoly as a natural phenomenon in a competitive framework rather than a state of affairs incompatible with a state of competition.[17] Perhaps because economic theory was faltering, or, perhaps because the mere fact that you dislike the outcome of a transaction is not a sufficient reason to ban it, in 1982 the Supreme Court declared unconstitutional a wave of restrictive anti-takeover laws that were passed by some states' legislatures.[18]

As mentioned above, to prevent monopoly power from being seized is not the only justification for supporting takeover regulation. We need now to analyse the other justification brought forward to maintain strict control over a takeover: namely, that it might injure shareholders, and more generally the functioning of the capital market.

The debate on this point is animated and brings us back to corporate governance.[19] Empirical evidence (starting with Jensen and Ruback 1984) tends to show that target's shareholders gain from takeovers, while the shareholders of the acquiring company, on average, neither gain nor lose from such an operation. Also, some empirical studies have looked at the evolution of productivity after a takeover and have found no significant effect, while other studies support the thesis that mergers increase companies' profitability and are correlated with growth (see Healey et al. 1992; Lichtenberger 1992). Putting aside empirical evidence – evidence that remains, anyway, unclear[20] – through what mechanism could takeovers increase, or lessen, the well-being of shareholders?

Again, to address this question we should recall the nature of the title owned by a shareholder, and in particular the limited liability property of that title. If, as argued in previous sections, limited liability is a way to acknowledge the control problem, that is, if it is essentially a compensation granted in order to surrender, at least in part, the control over the management of the

asset, then shareholders can hardly expect to have the company's assets managed exactly as it pleases them.

In that spirit, and as forcefully argued by Henry Manne, takeovers could even be seen as a blessing for shareholders. A shareholder with limited liability, and little incentive or capacity to exercise thorough control of the management could rely on outsiders to check on the quality of the work done by incumbent managers. If the latter perform poorly, if they do not get the best out of the company's assets, outside entrepreneurs will then attempt to seize the profit opportunity and shareholders will be able to sell their shares at a premium (Manne 1965).

This is surely a powerful argument. It is in fact the traditional argument used to underline the benefits from a property rights system: if you believe that a given asset can be managed in a more profitable way (in the widest sense of the word, that is, it might be that you attach a high but very subjective and unusual value to the asset), and if you can make the actual owner of the asset better off, then a transfer of property can occur that will make everyone better off.

In the case of a takeover, however, two complications may arise that need to be dealt with. First, it is a case of co-ownership (and therefore some shareholders may hold on to their shares, refusing to tender them to the acquirer). Second, one might argue that actual owners do not realize where their own interest lies. They might run after short-term benefits and, if everyone does the same, no company will be able to engage in strategies that are beneficial only in the long run. Let us start with that second problem: Can the short-sightedness and lack of information of investors force managers to adopt a short-sighted strategy?

The mere fact that some shareholders are interested only in the short-term benefit does not provide a sustainable argument against takeover. For, indeed, two cases may arise. Either, prior to the offer, the share was 'correctly' valued – by that we mean here that market participants had correctly assessed the quality of the work done by managers[21] – or the share was undervalued. In the first case, shareholders are surely not harmed by an offer to tender their shares at a premium. In the second case one must again distinguish between two possibilities: either the shareholders were aware of such undervaluation, or they were not. If they were, then they should rationally have bought more shares and hence their benefits from and chances of success of a takeover would be even greater. If they had not yet realized the quality of the work done by the incumbent management, then one may reasonably believe that the high premium offered at the takeover is based on the present value of the asset, given that good investments have been realized. Hence, through a takeover process, the shareholders will, at worst, have to share part of the profits that the 'raider', and not they, has discovered. This again can hardly be seen as an injury.

Now, it is still possible that some shareholders judge the offered premium to be too low compared to what they believe to be the real value of the share. In such circumstances, shareholders may be disappointed by the fact that other co-owners tender their shares. In other words, they may wish the takeover attempt to fail and be disappointed in their expectations. Conversely, they may wish the takeover to be successful – thinking that it is a good deal – and be disappointed by the behaviour of those shareholders who, holding on to their shares, force the acquirer to withdraw his/her offer. Even if understandable, such a disappointment by definition concerns a minority of shareholders and, anyway, this problem is not specific to takeovers; it is a general problem with co-ownership and arises as well in a proxy contest or any vote of the general assembly of the corporation. Hence, if the possibility of holding out is a real problem, it will, in our view, best be dealt with by shareholders themselves when drafting the chart of the company and defining the contractual arrangement binding all shareholders.[22] Now, due to the diversity of shareholders' preferences, it is to be expected that choices will differ from one company to another. For that same reason, it is difficult to find a clear interpretation to the change in share value that occurs when the board of a company decides to move to a state where corporate law makes takeover transactions more difficult (such as Delaware).[23] Indeed, some investors might take that move as good news because they trust their management and think they can control their behaviour through other means. But the reverse is also possible: shareholders can interpret the change as bad news because they wish to rely more on outside control. What one should worry about, in that case as in similar ones, is harmonization of the law across states. Such harmonization would reduce the alternatives open to shareholders as well as to managers. As long as contractual freedom is preserved, there is no reason to be, as a matter of principle, for or against anti-takeover strategies.

The above presentation stresses the diversity of ways of controlling the performance of the managerial team as well as the various possible motives behind a takeover. The bottom line is that such transactions can be profitable for all the parties in the long run, or to the disadvantage of one party or even both. Can we know beforehand what will be the case? While partisans of the rule of reason tend to answer positively, we have argued to the contrary that, the market being essentially a discovery procedure, a social costs–social benefits analysis of the consequences of a takeover is impossible, even to the best experts. In short, the rule of reason does not appear to us reasonable when one takes into account its cognitive requirement.

But even if it were possible to assess beforehand with 'enough' accuracy the outcome of the transaction, would that constitute a sufficient reason to prohibit it? To answer positively would be tantamount to greatly reducing

contractual freedom, for, if the idea is accepted in the case of takeovers, why not interfere as well in many, otherwise valid, property transactions?

Public choice considerations

The analysis so far has not enabled us to find clear arguments – economic or other – in favour of a regulation of takeover and insider trading; at least nothing that justifies the kind of strong prohibition that has developed. If one wishes to understand the origin of such regulations and rules one must therefore look in another direction and ask who, if anyone, benefits from those regulations. It is with this public choice approach of the problems that we shall end this chapter.

To start with, let us see who benefits sufficiently from a prohibition of insider trading to formulate a demand? The motivation of shareholders does not seem to be strong enough, since, as was argued earlier, a single shareholder does not lose much through insider trading and, furthermore, getting all shareholders together to lobby in favour of such a regulation is costly. If not from shareholders, the demand for regulation could eventually come from those managers who fear that insider trading transactions conducted by their colleagues will in the long run affect their reputation and salaries. Such managers may welcome a regulation that will impose control for them. But, it could also come from those managers who feel under threat from a hostile takeover, since insider trading may lower the cost of such an operation.[24] More importantly, and as suggested by Haddock and Macey, the demand for regulation could come from market professionals (see Haddock and Macey 1987; Macey 1991). Indeed market makers will always be on the wrong side of the transaction when dealing with an insider. Consequently, they will increase the bid–ask spread to protect themselves, thereby indirectly hurting professional investors. Finally, market analysts would be more than happy to see their main competitors, the few who are better informed than them, severely punished by the administration.

On the supply side, regulating agencies are interested in an enlargement of their mission and in the increase of power that goes with it.[25] In this respect, insider trading, because it is presented as a symbol of the war against supposedly unfair trading, as a battle for the weak against the privileged, and more generally as a necessary step towards the reconciliation of market and morality, is a benediction for regulators.

As far as takeovers are concerned, the demand for regulation could come from managers to the extent that the threat of a takeover is not always welcomed. That explanation, however, is hard to reconcile with the fact that most takeovers are friendly. Besides protecting managers' interest, the regulation could be welcomed by companies trying to stop the growth of their competitors.[26] Let us also note that competition law cases make extensive use

of experts (economists, lawyers and so on). Those experts will surely push the demand for regulation, or at least are unlikely to ask for a relaxing of the regulation. Finally, mention should be made of stakeholders' interest, and in particular of employees' motivation. As Bittlingmayer (2000, p. 743) argues, one of the reasons behind a hostile merger would be to allow for the kind of heavy restructuring which is so difficult to undertake from inside. One could then understand the hostility of stakeholders – in particular of employees – towards such a change of control.

On the supply side, the motivation would be identical to that already mentioned for insider trading. More regulation translates into more power for 'market authorities'. In many countries, the trend clearly has gone in that direction, and the level of fines that can be imposed gives a rapid idea of that new power. In a case of insider trading, the SEC can ask the Justice Department to initiate criminal prosecution (the fine can be as high as $2.5 million and up to ten years in jail). In addition, the SEC may pursue civil penalties seeking a permanent or temporary injunction, or seeking disgorgement of the profits realized, or decide on a monetary penalty that can be up to three times the profit gained or loss avoided. It is even allowed to pay a bounty to informers of up to 10 per cent of the penalty it has collected, not to mention other sanctions such as revocation, or suspension of activities for a broker.

Concluding remarks

Both insider trading and takeovers are the subject of a great many legal rules and regulations set at various levels (the Federation or the Union, the states, corporate law). In both cases the interference is motivated by the desire to protect the weak party to the transaction – because that party is believed to be uninformed or ill-informed – and hence to promote general (social) efficiency (including the well-being of third parties not directly involved in the transaction). We have argued that such regulations and rules can be called into question for various reasons. The most important one, in our opinion, is that this trend neglects the main features of that form of ownership, namely, shared ownership and limited liability. Because of those features, the words of the Massachusetts Supreme Judicial Court are more than ever to be remembered: '[Law] cannot undertake to put all parties to every contract on an equality as to knowledge, experience, skill and shrewdness. It cannot undertake to relieve against hard bargains made between competent parties without fraud'.[27]

Surely, institutional arrangements need to evolve. And, in that field as elsewhere, competition remains the safest route, albeit not a perfect one. Regulation from the top and harmonization of regulation across jurisdictions is what we should be afraid of.

Notes

1. According to Meulbroek (1992), 79 per cent of insider trading cases initiated by the Securities Exchange Commission between 1980 and 1989 were related to takeovers.
2. For a history of and extensive references on limited liability, see Carney (2000).
3. That solution required the development of the legal concept of a corporate body.
4. At this point, it would be interesting to investigate to what extent bankruptcy law offers some protection to third parties, thereby reshaping the control problem. However, a discussion of bankruptcy law is beyond the limit of this chapter.
5. Of course, it is still possible to argue that, without insider trading, capitalization would have been even higher.
6. On the effectiveness of insider trading regulations see, for instance, Boardman et al. (1998) who are rather optimistic, and Jaffe (1974) or Seyhun (1992) who are sceptical.
7. Rule 16b also relates indirectly to insider trading.
8. The case, furthermore, was one of tipping, that is, the transaction was not conducted by an insider. Instead, it was conducted by an outsider who had been informed (tipped) by an officer of the company which shares were traded.
9. *Chiarella v. United States*, 445 US 222 (1980) and *Dirks v. SEC*, 463 US 646 (1983).
10. Meanwhile, in 1989, a European Community directive (Council Directive 89/592) harmonized insider trading regulations throughout the Community.
11. Other ways, such as the election of a new board of directors through a proxy contest, are usually more expensive.
12. This is actually the Supreme Court's interpretation of the law. We come back to the public choice aspect of regulation in the last section. For a history of takeover regulation see Kovacic and Shapiro (2000) for the United States, or Neumann (2001) for Europe.
13. More precisely, the Act stipulates that, following a tender offer, shareholders have a right to withdraw their offer to tender during the first seven days. The Act also imposes prorationing of tendered shares after 60 days and stipulates that increased offers apply retroactively to shares tendered earlier. It also requires from any acquirer who buys 5 per cent of outstanding shares on the open market that he/she discloses his/her purchase to market authorities.
14. Brown shoe case, reported in Glais (1992).
15. That those conglomerates made little economic sense can be deduced from the fact that, as soon as horizontal mergers were allowed to resume, the '*fashion*' changed, many of those conglomerates being dismantled and the new dogma becoming one of coherent development centred on a clearly identified know-how. Bittlingmayer (1996, 2000) goes further and argues that a negative correlation could exist between the degree of takeover prohibition and the return on stock exchanges. In particular, the crisis and crashes that took place in the twentieth century would have followed announcements of tighter regulations.
16. Two publications, by Posner (1976) and Bork (1978), have revisited the case for antitrust.
17. For more details on what is meant here by 'a discovery process', see Hayek (1978) or Kirzner (1997). The traditional, neoclassical approach still has many adherents. See, for instance, Neumann (2001) for a recent presentation.
18. In *Edgar v. Mite*, 457 US 624 (1982). We must note, however, that a later wave of anti-takeover law more carefully drafted by the states' legislations was upheld.
19. The corporate governance side of the debate has various dimensions according to the structure of ownership in the company, and more generally in the jurisdiction. Hence, in Germany or Japan, because banks hold large shares of companies, attempts to gain managerial control through takeovers are rare. In the United States, on the contrary, the Pujo 'Money trust investigation' drove banks off the boards of directors in 1912, and investment funds and other financial intermediaries are also restricted in their stockholdings, giving corporate governance a different outlook.
20. But recall that any voluntary transaction, even in a strict Paretian–neoclassical framework, increases the well-being of both parties. Although of course some will argue that this is untrue when there are informational problems.

21. Of course, only as the future unfolds will it be possible to find out whether or not the share was correctly valued. Market evaluation always rests on subjective expectations.
22. A central point in the actual drafting of the European Union takeover code is precisely the harmonization of the rules concerning voting rights. But should not this be a matter left to the company?
23. Empirical studies show that this change is not significant. See Bittlingmayer (2000, p. 740).
24. Through so-called 'warehousing', the raider informs outsiders of his/her intention to take over the target, so as to increase the chances to have enough shares tendered once the operation is publicly launched. Rule 14e-3 is mainly directed against this type of strategy.
25. To use Niskanen's terminology, they wish to maximize 'the size of their office'. See Niskanen (1971).
26. See McChesney and Shughart (1995) for evidence that antitrust is most often a weapon for rent seekers to wield against their rivals.
27. 186 N.E. 660 (Mass. 1933), as quoted in Bainbridge (2001, p. 8).

References

Bainbridge, Stephen M. (2001), 'The Law and economics of insider trading: A comprehensive primer', <http://papers.ssrn.com/sol3/papers.cfm?abstract.id=261277> (downloaded 23 January 2004).

Bittlingmayer, G. (1996), 'Merger as a form of investment', *Kyklos*, **49**, 127–53.

Bittlingmayer, G. (2000), 'The market for corporate control (including takeover)', in Bouckaert and De Geest (eds), Vol. III, 5640, pp. 725–71.

Boardman, A., Z. Stuart Liu, M. Sarnat and I. Vertinsky (1998), 'The effectiveness of tightening illegal insider trading regulation: the case of corporate takeovers', *Applied Financial Economics*, **8**, 519–31.

Bork, Robert (1978), *The Antitrust Paradox – A Policy at War With Itself*, 1993 edn, New York: Free Press.

Bouckaert, Boudewijn and Gerrit De Geest (2000), *Encyclopedia of Law and Economics*, 5 vols, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Carney, W. (2000), 'Limited liability', in Bouckaert and De Geest (eds), pp. 659–91.

Fama, Eugene (1970), 'Efficient capital markets: A review of theory and empirical work', *Journal of Finance*, **25**, 383–417.

Glais, Michel (1992), *Economie Industrielle* [Industrial Organization], Paris: Litec.

Haddock, D.D. and J.R. Macey (1987), 'Regulation on demand', *Journal of Law and Economics*, **30**, 311–52.

Hayek, F.A. (1945), 'The use of knowledge in society', *American Economic Review*, **35** (4), 519–30.

Hayek, F.A. (1978), 'Competition as a discovery procedure', in Friedrich A. Hayek, *New Studies in Philosophy, Politics, Economics and the History of Ideas*, Chicago: University of Chicago Press, pp. 179–90.

Healey, P.M., Krishna G. Palepu and Richard S. Ruback (1992), 'Does corporate performance increase after mergers?', *Journal of Financial Economics*, **31**, 135–75.

Jaffe, J.F. (1974), 'The effect of regulation changes on insider trading', *Bell Journal of Economics and Management Science*, **5**, 93–121.

Jensen, M.C. and R.S. Ruback (1983), 'The market for corporate control: the scientific evidence', *Journal of Finance*, **11**, 5–55.

Kirzner, I. (1997), 'Entrepreneurial discovery and the competitive process: an Austrian approach', *Journal of Economic Literature*, **35**, 60–85.

Kovacic, W.E. and C. Shapiro (2000), 'Antitrust policy: A century of economic and legal thinking', *Journal of Economic Perspectives*, **14** (1), 43–60.

Lichtenberger, Frank R. (1992), *Corporate Takeovers and Productivity*, Cambridge, MA: MIT Press.

Macey, Jonathan R. (1991), *Insider Trading: Economics, Politics and Policy*, Lanham, MD: AEI Press.

Manne, H.G. (1965), 'Mergers and the market for corporate control', *Journal of Political Economy*, **73**, 110–20.
Manne, Henry G. (1966), *Insider Trading and the Stock Market*, New York: Free Press.
McChesney, Fred S. and William F. Shughart II (eds) (1995), *The Causes and Consequences of Antitrust: The Public Choice Perspective*, Chicago, IL: University of Chicago Press
Meulbroek, L.K. (1992), 'An empirical analysis of illegal insider trading', *Journal of Finance*, **47**, 1661–9.
Neumann, Manfred (2001), *Competition Policy – History, Theory and Practice*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Niskanen, William A. (1971), *Bureaucracy and Representative Government*, Chicago, IL: Aldine-Atherton.
Posner, Richard (1976), *Antitrust Law: An Economic Perspective*, Chicago, IL: University of Chicago Press.
Seyhun, H.N. (1992), 'The effectiveness of the insider-trading sanctions', *Journal of Law and Economics*, **35**, 149–82.

17 The new property rights theory of the firm

*Pierre Garrouste**

Introduction

The idea that the firm can be conceived on the basis of the definition and distribution of property rights has generated an important literature. Coase (1960) is one of the first who emphasized the idea that property rights are effective in economics. Alchian (1965) and Demsetz (1967) clarify the notion of property rights and extend its application in economics. Grossman and Hart (1986) define the firm with an explicit reference to the distribution of ownership of the assets, while Hart and Moore (1990) give a perfect formal presentation of a property rights-based theory of the firm (with only two parties and without looking at the internal organization of the firm, see below). They 'identify the firm with the assets it possesses and take the position that ownership confers residual rights of control over the firm's assets: the right to decide how these assets are to be used except to the extent that particular usages have been specified in an initial contract' (ibid., p. 1120).[1]

The main problem that the new property rights theory of the firm *à la* Grossman and Hart (1986) tries to solve, concerns the effect that ownership of assets has on the incentives of two parties (usually a buyer and a seller) to invest *ex ante* in non-contractible assets,[2] knowing that they share *ex post* the quasi-rents that their investments produce. The two parties have the possibility of trading outside, that is to say, with a third party.

We shall present some of the main points arising from this conception of the firm and show that they can be and have been challenged. The first important point, examined in the next section, concerns the definition of property rights and the possibility of identifying two different property rights theories; such authors as Alchian, Demsetz and Barzel support the first one, which is usually called 'old property rights theory' (OPRT), authors like Grossman, Hart and Moore develop the second ('new property rights theory', that is, NPRT). This section also deals with the relations between property rights and decision rights and with the possibility of building up a theory of the firm on the idea that the owner of an asset is the 'residual claimant' when it is impossible to write complete contracts.

The second point concerns the specificity of NPRT as compared to other similar approaches, such as the transaction cost theory of the firm (TCT) and the incomplete contracts theory (ICT).[3] Following Gibbons (2002) and

Whinston (2003) we shall show that the idea that NPRT is a formal version of TCT is unfounded and that NPRT and ICT are often confused.

The third point concerns the possibility of defining an optimal distribution of ownership, that is to say the best way to allocate the property rights in terms of the incentives to invest (in human assets) that this allocation implies. The subsequent section presents the NPRT analysis of the size of the firm on the basis of the distribution of property rights and the difficulty it has in dealing with the problem of the organizational structure of the firm, even if some recent contributions are handling this problem. Next, we analyse the way that NPRT solves the bargaining and renegotiation problems, and then we present the empirical evidence in favor of NPRT. The penultimate section exposes some pending problems of NPRT and the final section concludes.

The two property rights theories

The definition of property rights

The way property rights are conceived in property rights theory is not unique. Demsetz (1967) considers that one of the main 'function[s] of property rights is that of guiding incentives to achieve a greater internalization of externalities' (ibid., p. 348). He considers that new property rights emerge in order to answer new problems linked with 'the desires of the interacting persons for adjustment to new benefit–cost possibilities' (ibid., p. 350). He then analyses the relations between property rights and ownership. He develops the idea that private ownership internalizes many of the external costs that collective ownership induces. As an example, private ownership reduces negotiation costs because the owner negotiates directly with those who want to use his/her property whereas all the members of the community negotiate together the use of collective property in the presence of collective ownership.

Following Alchian, Barzel (2001, p. 12) distinguishes economic and legal rights. He assumes that 'economic rights reflect the ability (in expected terms) to benefit from a good (or a service)', when 'legal rights are the rights that the state recognises as those of a particular individual or a set of individuals' (ibid., p. 14). Barzel shows that those two kinds of rights need to be distinguished because individuals attempt to maximize their economic rights subject to the others' attempt to maximize their own economic rights and the distribution of more or less perfectly enforced legal rights. On this basis Barzel defines the firm as 'a nexus of the agreements and parts of agreements guaranteed by centralised equity capital and enforced without the state's assistance. The scope of the firm is the ratio of its guaranteeing capital to that of (some measure of) its expected guarantee payments' (ibid., p. 21).

The literature based on the conceptions presented above is usually referred as 'old property rights theory'.

In Grossman and Hart (1986) and Hart and Moore (1990) the problem of the nature of the firm is based on a precise and specific definition of property rights. They write that the 'property rights approach takes the point of view that the possession of control rights is crucial for the integration decision' (Hart and Moore 1990, p. 1121 note 3). The rights possessed by the owner of an asset are only due to his/her ability to exclude others from using this asset. It is an interesting way of dealing with property rights; even if the basic approach needs to be completed because it follows that the problem of delegating decisions on the use of the assets can be analysed (see below, and Hart and Hölmstrom 2002). Grossman and Hart (1986) and the following literature (which is generally regarded as the 'property rights theory of the firm' – see Williamson 2000, p. 605) – and that we present principally in this chapter) is often called 'new property rights theory'. As we shall see below, it stresses the distinction between the non-residual rights of control, that is, the rights that can be enforced *ex ante* in an initial contract, and the residual rights of control that cannot.[4]

What distinguishes the Grossman, Hart and Moore (NPRT) and the Alchian, Demsetz and Barzel (OPRT) way of dealing with the property rights problem is not only that the former develops formal models, but also (i) that NPRT considers that the distribution of ownership conditions the investments (in human assets) of the economic agents, whereas OPRT assume that economic agents are able to privately enforce agreements; (ii) that according to NPRT, even if the ownership of the stream of profit is 'in practice' linked with the ownership of control, it is not necessarily so; whereas for OPRT the property rights are an economic notion, in the sense that they are directly linked with the capacity to get some profits; and finally (iii) that according to OPRT, ownership and control need to be distinguished whereas NPRT assumes an identity between ownership and control (this assumption has recently been revised, see below).

We shall focus attention on NPRT because of the importance of the literature it has provoked and the sophistication of its developments.

The residual claimant and the residual rights of control

The idea that those who own the assets have the rights to obtain the benefits that are not initially defined by a contract (that is, the residual benefits) is a notion that is not accepted by all the advocates of the property rights conception of the firm. As an example, following Grossman and Hart (1986), Hart and Moore (1990, p. 1121, note 3) distinguish, 'between ownership in the sense of possession of the residual control rights over assets and ownership in the sense of entitlement to an asset's (verifiable) profit stream'. In Grossman and Hart (1986) ownership and control are not distinguished, as they are in Demsetz (1967) or Pejovich (1990), and the problem of the nature and scope

of the firm is analysed on the basis of the allocation of the *residual rights of control*. The starting point is to consider that 'contractual rights can be of two types: specific rights and residual rights' (Grossman and Hart 1986, p. 692). The former are defined in the initial contract, but not the latter, due to exceedingly high costs.[5]

In fact NPRT does not deal with the problem of sharing a stream of profit, as in Barzel. The problem is to show how the ownership of the assets induces some effects on the incentives of the parties to invest in human assets. As Maskin and Tirole (1999a, p. 142) write,

> [T]he property-rights literature assumes that the only feasible contracts are *unconditional* ownership contracts; namely, the date-0 contract specifies that the physical asset belongs to the seller, the buyer, or to both (in which case neither can use the physical asset to trade with an alternative partner without the other's consent). The good to be trade is determined *ex post* through bargaining.

The important point is the idea that the ownership contracts are unconditional, which implies that the property rights are not dispersed between many owners. This avoids resulting problems, for example, the governance of firm one.

NPRT, TCT and ICT

It is often claimed that NPRT is a formalization of TCT.[6] It is also usually held that NPRT and ICT are identical. We shall clarify these two points.

As far as the relations between TCT and NPRT are concerned, Williamson (1996) saw NPRT as a good formalization of the TCT arguments. Hart and Hölmstrom (2002) write that the NPRT literature is built on the earlier TCT of Williamson. This is due to the similarities between the two theories. They start 'from the same premises of incomplete contracts and *ex-post* quasi-rents' (Whinston 2003, p. 4). The difference between TCT and NPRT is then often reduced to a methodological one. Although TCT is not formalized, NPRT is. The direct consequence of this is that all the empirical literature supporting TCT also supports NPRT. This idea, however, has been challenged by Williamson (2000), Whinston (2003) and Gibbons (2002).

According to Williamson

> [The] most consequential difference between the TCE [transaction cost economics] and the GHM [Grossman, Hart and Moore, that is, NPRT] setups is that the former holds that maladaptation in the contract execution interval is the principal source of inefficiency, whereas GHM vaporize *ex post* maladaptation by their assumptions of common knowledge and costless *ex post* bargaining. The upshot is that all the inefficiency in GHM is concentrated in the *ex ante* investments in human assets (which are conditional to the ownership of physical assets). (Williamson, 2000, p. 605)

According to Whinston, there are two main differences (plus the methodological gap indicated above). The first is that although NPRT focuses on *ex ante* investment distortions, TCT is much more interested in the *ex post* haggling and maladaptation. Second, and in contrast with TCT, which considers 'that opportunism could be mitigated by bringing the transaction within the firm (with resulting bureaucratic costs), NPRT assumes that this hazard is present in all modes. It is so because investment and trading decisions remain fundamentally decentralized in NPRT, regardless of the structure of asset ownership' (Whinston 2003, p. 4). If ownership influences the incentives in NPRT, it has no effect on the coordination of the investments. Indeed, in a simple NPRT model (one buyer, one seller), ownership influences the decisions to invest for the parties but it has no effect on investment coordination. In order to stress those differences, Whinston uses a simple property rights model (a linear-quadratic one). This model shows that the parameters that are supposed to explain the choice between integrating and not integrating do not have the same effects as those predicted by TCT and even by Hart (1995). Gibbons expresses the same idea, although he considers the TCT as a rent-seeking theory. He writes that 'in property rights, it is the *ex ante* integration decision that determines *ex ante* incentives and hence total surplus, whereas in the rent seeking theory it is the *ex post* integration decision that determines *ex post* haggling and hence total surplus' (Gibbons 2002, p. 4).

Eventually TCT assumes bounded rationality, whereas according to NPRT, agents are Savagian – that is, perfectly rational. We shall analyse the consequences of this last assumption in the penultimate section.

The comparison between NPRT and ICT is much more difficult to analyse. In fact those two theories are often confused and considered to be identical. Brousseau and Glachant (2002, p. 10, emphasis added) defend this opinion when they write that 'ICT thus came to examine the impacts of the institutional framework on contract design, *though its roots lay in the study of property-rights allocations on the distribution of the residual surplus between agents and on their incentives to invest*'. This vision is based on the idea that it is possible to identify different generations of models in ICT. As an example, Brousseau and Fares (2000) show that Grossman and Hart (1986) and Hart and Moore (1988) differ in the way they deal with the rationality assumption. In this way, NPRT and ICT are supposed to belong to the same research program but can be distinguished only on the basis of certain assumptions.

In order to clarify the relations between NPRT and ICT we shall use the arguments developed by Maskin and Tirole (1999a). They consider that three assumptions are at the basis of NPRT[7] and that it is not a problem of 'unforeseeability of future contingencies, which is often stressed in the literature' (ibid., p. 140). They consider that '*the three assumptions driving the*

conclusion that property rights suffice are i) parties' ability to renegotiate, ii) the equivalence of outside opportunities to self-consumption of the asset, and iii) parties' risk-neutrality' (ibid., original italics). As we shall see below, Maskin and Tirole (1999a) challenge one of the main results of Hart and Moore (1990) concerning the possibility of determining an optimal distribution of ownership. But an important point to stress is that if NPRT is not based on incomplete contracting, then the criticisms raised by Maskin and Tirole (1999b) are not against NPRT – in fact, they discuss the relevance of ICT. Then, if NPRT is not rooted in ICT, the former is not vulnerable to the attacks made against the latter. However, if the NPRT is using ICT as a tool then the shortcomings of ICT are also shortcomings of NPRT.

This way of clarifying the difference between ICT and NPRT suggests that the scope of problems the NPRT tries to deal with is much more important in comparison with ICT; and that ICT can then be seen as a 'tool' for NPRT.

To sum up, it would seem that NPRT is not a formal version of TCT even if they are both linked with ICT;[8] and the idea that NPRT and ICT are identical is open to discussion. It is clear that NPRT and ICT share some assumptions but it is not self-evident that NPRT and ICT are solving the same problems. The main problem NPRT wants to solve is the way one can define optimal ownership, a problem that is not central to the ICT approach. Indeed ICT is trying to solve the problem of the existence of a possible contract when uncertainty is the case and when it is impossible for a third party to look at the way the two parties are behaving without looking at its consequences in terms of ownership distribution.

The optimal distribution of ownership

Grossman and Hart (1986) as well as Hart and Moore (1990) define what they consider to be the main problem that the property rights theory of the firm has to deal with. As a result, they determine the best way to assign property rights in order to solve the Coase problem of the nature and boundaries of the firm. It is a normative conception of the problem of the nature and scope of the firm. Their model is based on six main assumptions. The first two are well known and define the cost and return functions. Assumption three says that if an individual is not a member of a coalition, his/her contribution is nil. The fourth one considers that there is a complementarity at the margin between individual investments of those who are in the same coalition. The fifth is a superadditivity assumption and the sixth assumes that the marginal return on investment is an increasing function of both the number and the assets of a coalition (Hart and Moore 1990, p. 1127). Finally they assume that individuals are bargaining according to the Shapley value.[9] They obtain some important and general results concerning (i) the way it is necessary to define the ownership of the assets (joint ownership or one or the other

party ownership) and (ii) the optimality of this definition, which depends on the investment of the parties and the characteristics of the assets. As an example, joint ownership is suboptimal because it produces low incentives for the parties to invest in human assets.

This result has been challenged by Maskin and Tirole (1999a). They adopt a Nash bargaining solution and by using an NPRT model they show that, if the cost incurred by the seller is high when the buyer trades outside (that is with another party) and if the gain for the buyer is low when the seller trades outside, there is no outside opportunity and it is impossible to do better than joint ownership. They then bring in an option to sell and show that this possibility deters underinvestment. Finally, they introduce the possibility of renegotiation and collusion. They also obtain a result different from the Hart and Moore (1990) idea that join ownership is not efficient. They stress, however, that their results are not denying NPRT. They only consider their results as complements and improvements of the NPRT fundamental results.

Property rights, the scope of the firm and its organizational structure

Until recently NPRT was silent on the problem of the scope of the firm (Hölmstrom and Roberts 1998). For example, Gibbons asks 'where are the managers?' (2002, p. 5). According to Hart and Hölmstrom (2002) if NPRT allows the determination of the boundaries of the firm, this applies to the owner-managed firms, but not to large companies. In order to solve this difficulty they propose an extension of the NPRT models. They first define the unit of production as the basic element that corresponds to an activity. It is owned by a boss and has a manager and possibly workers. The unit of production can be managed by the boss or a manager. The key modification of the NPRT basic models is the assumption that 'decisions can be transferred only through ownership, and are not even *ex post* contractible' (ibid., p. 2). They then assume that 'each unit generates two kinds of benefit: monetary profit and private (non transferable) benefits in the form of job satisfaction for those working in the unit' (ibid., p. 3). They then propose two models in order to answer not only the question of the incentives to invest *ex ante* but also of who invests. The first is a two-unit model based on a yes-or-no choice. The second introduces an outsourcing decision.

These models are interesting because they can be extended to the analysis of the hierarchy of the firm. Indeed, it seems possible to introduce delegation possibility as in Hart and Moore (2000). The main idea of this paper is that since it is costly for the owner of the assets to decide how to use them, he/she can delegate this decision. The only problem of this model of delegation is that the owner is not free to choose who is going to decide on the assets (as opposed to Aghion and Tirole 1997). In fact, there is a 'command chain' where the individuals are hierarchically localized and their place on this

chain defines the right they have to decide. An interesting result is that the probability of the individual having an idea on the use of an asset is negatively correlated with his/her right to decide how to use the asset. A much more general analysis of the problem of the hierarchy is given by Rajan and Zingales (2001), who propose a model where the difficulties of defining property rights on the assets are taken into account. In fact, *ceteris paribus*, the more the asset can be protected by property rights, the deeper is the vertical hierarchy. In contrast, if it is difficult to define property rights on the assets (the critical resources), then the hierarchy needs to be flat.

All these recent works show that it would be possible for the NPRT to deal with the organizational structure of the firm.

Bargaining and renegotiation

Bargaining is an important problem when one wants to present and appraise NPRT results. NPRT assumes that *ex post* the parties bargain in order to share the surplus (that is the quasi-rents that their investments yield). NPRT models usually assume that ownership has no effect on efficient bargaining. In fact both Grossman and Hart (1986) and Hart and Moore (1990) essentially analyse the effect of ownership on the incentives of the parties to invest. In particular, ownership modifies the parties' disagreement payoffs (that is, the payoffs they obtain if they trade outside). When modeling bargaining, the 'traditional' NPRT literature often uses either the Shapley value (Hart and Moore 1990) solution or the Nash one (Maskin and Tirole 1999a). Recently, however, the NPRT literature has introduced the possibility that 'by altering the disagreement payoffs, ownership changes can have significant effects on how efficiently managers bargain with each other in the presence of private information' (Matouschek 2002, p. 2). According to Matouschek, four ownership structures can be optimal: buyer integration, seller integration, non-integration, and joint ownership (in contrast with Hart and Moore 1990 and in line with Maskin and Tirole 1999a). In fact, Matouschek shows that when two parties are bargaining in order to share the quasi-rents in a context of private information, it is optimal for those parties to minimize the aggregate disagreement payoff when, for a given distribution of trade payoffs, the minimum expected quasi-rents are important. If this is not the case, their optimal solution is to maximize their aggregate disagreement payoff (see Matouschek 2002, p. 28).

In the initial NPRT literature, renegotiation was not taken into account. There is no reference concerning this point in Grossman and Hart (1986), Hart and Moore (1990) or Hart (1995). On the other hand, within the ICT context the possibility of renegotiating is an important topic (see Hart and Moore 1988; Maskin and Moore 1999). The explanation is that there is no problem of renegotiation in NPRT: if the problem is to define the optimal

ownership solution, and if ownership is equivalent to control and if bargaining is efficient, then renegotiation is not relevant. This position can be challenged. First it could be empirically justified to separate ownership from control (see the following section). Furthermore, Maskin and Tirole (1999a) introduce this possibility on a theoretical basis. They develop the idea that the buyer's and the seller's incentives are determined by their marginal value and marginal cost, respectively. Such incentives generate the respective payoffs of the two parties. They show that renegotiation, even if it does not give strong predictions in terms of ownership, can be introduced in an NPRT model.

The empirical evidence of the property rights theory

Until recently there were no direct empirical tests of NPRT. As presented above, the idea was that direct empirical tests in favor of TCT[10] were taken as indirect empirical confirmation of NPRT, since NPRT was considered to be the formal version of TCT. This viewpoint has now been abandoned (see above).

Hölmstrom and Roberts (1998) give another explanation for the lack of direct empirical support in favor of NPRT. They point out that according to NPRT, the *level* of asset specificity does not influence the allocation of ownership because investments are only driven by *marginal* returns. They add: 'this is problematic for empirical work, partly because margins are hard to observe when there are no prices and partly because some of the key margins relate to returns from hypothetical investments that in equilibrium are never made' (ibid., p. 79).

However, some direct empirical tests of NPRT are now available. Baker and Hubbard (2003, p. 31) answer the question about 'what determines who owns the assets' by looking at the US trucking industry. They find evidence that contractibility influences ownership. More precisely, the introduction of on-board computers improves the possibility of contracting and also leads to more integrated asset ownership, that is to say more integration. Elfenbein and Lerner (2003) evaluate the effects on ownership of the control rights in the internet alliances. Analysing 100 contracts they show that there is strong evidence in favor of Grossman and Hart (1986). However, there is also a sensitivity of control rights to the bargaining power of the parties. This result is not in line with the 'traditional' NPRT models and it is interesting inasmuch as it stresses the idea that ownership and control are not necessarily the same thing. Indeed the NPRT literature always considers that ownership and control are confused. This is disputable *per se*, but the diffusion of technologies that are based more and more on non-contractible assets (human assets) is making this problem very real. In fact, the new technologies of information and communication are making

the relations between ownership and control too complex to be handled by the initial NPRT models.

Pending problems

We shall present three main criticisms that could be made against the NPRT. The first is linked with the idea that there is no third party. As a consequence, there are no constraints on bargaining between the parties. The second is that the parties are supposed to be Savagian in terms of rationality. These two criticisms concern one of the NPRT assumptions about the existence of imperfect contracting (Whinston 2003). Indeed, ICT assumes perfect rationality, symmetric information and the idea of 'observability but non-verifiability' of the decisive variables. This last idea means that even if it is possible to observe the variables, a third party is not able to verify them; that is, no third party is able to know whether the two parties have behaved correctly, for he/she cannot verify this point. The idea that NPRT assumes perfect rationality of the parties indicates another difference between TCT and NPRT, in that according to TCT transaction costs also originate from the bounded rationality of the individuals[11] (see above).

This lack of a third party and the assumption that the two parties are perfectly rational are linked and the problem is then one of coherence. As pointed out by Brousseau and Glachant (2002), if the judge is not rational because he/she cannot verify the key variables, how is it possible to assume that the parties are? 'It would be more consistent to resort to a hypothesis of bounded rationality for all actors – the parties and the judge – as it is the case in the TCT' (ibid., p. 12).

Another shortcoming of NPRT (depending on its relationship with ICT) is the famous 'irrelevance theorem' due to Maskin and Tirole (1999b). Their argument is that 'if the parties have trouble foreseeing *physical* contingencies, they can write contracts that *ex ante* specify only the *payoff* contingencies' (Maskin and Tirole 1999b, p. 84). This irrelevance theorem shows in its first part (Theorem 1) that if there is no renegotiation the fact that the states of nature are not describable does not interfere with the possibility of setting up a welfare-neutral optimal contract, that is, a contract such that if two states of nature are payoff-equivalent it gives the same utilities in the two states. Maskin and Tirole then demonstrate (Theorem 2) that if 'the preferences satisfy the state-independence of the ratios of marginal utilities of money and effort is unidentified', then it is possible to define a welfare-neutral complete contract (ibid., p. 94). These two results are important because their basic assumptions are frequent in the ICT literature. Indeed, the idea that NPRT is not derived from ICT (as asserted in Maskin and Tirole (1999a), even if in note 3 this idea is seen as a potentiality) is important because it is not affected by the irrelevance theorem.

If, however, NPRT is rooted in ICT, then an answer is needed. Hart and Moore (1999) believe that there is a strong connection between the two because they answer the questions of Maskin and Tirole (1999b).

Conclusion

NPRT is one of the most interesting theories of the firm in economics. It undertakes to define the nature and the boundaries of the firm in terms of efficiency in ownership. The point is to select the best way to allocate ownership in order to generate incentives for two parties to invest in non- contractible assets (that is, human ones). This way of dealing with the choice between 'make or buy' is a challenge to the TCT approach and can be seen as not 'rooted' in ICT. The main important result of NPRT concerns the definition of optimal ownership, even if this point has been challenged. The second important result of NPRT is its recent emphasis on the internal organization of the firm. This is quite new and needs to be completed and developed. Finally, some recent empirical results seem to corroborate the NPRT fundamental assumption whereby ownership matters in economics. The essential problem that NPRT has to solve is to justify the fact that some variables cannot be verified even if observable.

Notes

* I thank Enrico Colombatto, Claude Ménard, Emmanuel Raynaud, Stéphane Saussier and Anne Yvrande-Billon for their comments. The usual caveats apply. I also thank the International Centre for Economic Research (ICER) for its financial support and the Department of Economics 'S. Cognetti de Martiis' of the University of Turin for its hospitality.

1. Barzel (1989) is also an impressive contribution to the economics of property rights. See also the reviews of this literature in Hart (1995) or Hölmstrom and Tirole (1989) and the special issue of the *Review of Economic Studies* (1999).
2. *Ex ante*, that is before the sharing of the quasi-rents. Non-contractibility of the assets is due to the fact that those assets are human.
3. We use the word 'theory' as a simplification. However, although it is easy and not confusing to speak of the general equilibrium theory, the use of the term 'theory' for these different approaches can be challenged. We do not discuss this problem in this chapter.
4. An interesting criticism of this distinction is given by Demsetz (1998).
5. See note 4.
6. 'One still sometimes hears mistaken views such as "Grossman and Hart (1986) formalised Williamson (1979)"' (Gibbons 2002, p. 2).
7. For them, the Hart and Moore (1990) model is no more than a bilateral monopoly model.
8. This is not to say that NPRT and TCT are also subsets of ICT. For a presentation of these different theories, see Brousseau and Glachant (2002).
9. 'The Shapley value gives agent *i* his expected contribution to a coalition, where the expectation is taken over all coalitions to which *i* might belong' (Hart and Moore 1990, p. 1129).
10. For a presentation of the empirical supports of TCT, see Masten and Saussier (2002).
11. Although it is not without problems, we shall not discuss this assumption of TCT here.

References

Aghion, P. and J. Tirole (1997), 'Formal and real authority in organizations', *Journal of Political Economy*, **105**, 1–27.

Alchian, A.A. (1965), 'Some economics of property rights', *Il Politico*, **30**, 816–29.

Baker, G.P. and T.N. Hubbard (2003), 'Contractibility and asset ownership: on-board computers and governance in U.S. trucking', mimeo, National Bureau of Economic Research, Washington, DC.

Barzel, Y. (1989), *The Economics of Property Rights*, Cambridge: Cambridge University Press.

Barzel, Y. (2001), 'A measurement cost based theory of the firm', mimeo, University of Berkeley.

Brousseau, E. and M. Fares (2000), 'The incomplete contract theory and the new-institutional economics approaches to contracts: substitutes or complements', in C. Ménard (ed.), *Institutions, Contracts and Organizations: Perspectives from New Institutional Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 399–421.

Brousseau, Eric and Jean-Michel Glachant (2002), *The Economics of Contracts, Theory and Applications*, Cambridge: Cambridge University Press.

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

Demsetz, H. (1967), 'Toward a theory of property rights', *American Economic Review*, **57** (2), 347–59.

Demsetz, H. (1998), 'Book Review of *Firms, Contracts and Financial Structure: Clarendon Lectures in Economics*, by Oliver Hart', *Journal of Political Economy*, **106** (2), 446–52.

Elfenbein, D. and J. Lerner (2003), 'Ownership and control rights in Internet portal alliances, 1995–1999', *Rand Journal of Economics*, **34** (2), 356–69.

Gibbons, R. (2002), 'Four formal(izable) theories of the firm', mimeo, MIT and NBER.

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), 691–718.

Hart, Oliver (1995), *Firms, Contracts, and Financial Structure*, Oxford: Clarendon.

Hart, O. and B. Hölmstrom (2002), 'A theory of firm scope', mimeo, MIT and NBER.

Hart, O. and J. Moore (1988) 'Incomplete contracts and renegotiation', *Econometrica*, **56** (4), 755–85.

Hart, O. and J. Moore (1990), 'Property rights and the nature of the firm', *Journal of Political Economy*, **98** (6), 1119–58.

Hart, O. and J. Moore (1999) 'Foundations of incomplete contracts', *Review of Economic Studies*, **66**, 115–38.

Hart, O. and J. Moore (2000), 'On the design of hierarchies: coordination versus specialization', mimeo, Harvard University and London School of Economics.

Hölmstrom, B. and J. Roberts (1998), 'The boundaries of the firm revisited', *Journal of Economic Perspectives*, **12** (4), 73–94.

Hölmstrom, B. and J. Tirole (1989), 'The theory of the firm', in Richard Schmalensee and Robert Willig (eds), *Handbook of Industrial Organization*, Amsterdam: North-Holland, pp. 61–133.

Maskin, E. and J. Moore (1999) 'Implementation and renegotiation', *Review of Economic Studies*, **66**, 39–56.

Maskin, E. and J. Tirole (1999a), 'Two remarks on the property-rights literature', *Review of Economic Studies*, **66**, 139–49.

Maskin, E. and J. Tirole (1999b), 'Unforeseen contingencies and incomplete contracts', *Review of Economic Studies*, **66**, 83–114.

Masten, S.E. and S. Saussier (2002), 'Econometrics of contracts: an assessment of developments in the empirical literature on contracting', in Brousseau and Glachant (eds), pp. 273–92.

Matouschek, N. (2002), 'Information and the optimal ownership structure of firms', Working Paper no. 0028, Center for the Study of Industrial Organization, Northwestern University.

Pejovich, S. (1990), *The Economics of Property Rights: Towards a Theory of Comparative Systems*, London: Kluwer Academic.

Rajan, R. and L. Zingales (2001), 'The firm as a dedicated hierarchy: a theory of the origins and growth of firms', *Quarterly Journal of Economics*, **116** (3), 805–51.

Whinston, M. (2003), 'On the transaction cost determinants of vertical integration', *Journal of Law, Economics and Organization*, **19** (1), 1–23.
Williamson, Oliver E. (1996), *The Mechanism of Governance*, Oxford: Oxford University Press.
Williamson, O.E. (2000), 'The new institutional economics: taking stock, looking ahead', *Journal of Economic Literature*, **38**, 595–613.

18 On intellectual property rights: patents versus free and open development

Alan G. Isaac and Walter G. Park

Introduction

Intellectual property rights (IPRs) encompass a broad array of legal protections, including patent rights, copyrights, trademark rights, plant breeders' rights, protection of trade secrets, industrial designs, layout designs for integrated circuits, and geographical indications (regarding the origin of goods and services). These legal protections serve diverse purposes and functions, and the institutions supporting them are also quite distinct. In this chapter we focus on the patent system.

Patents are a primary instrument by which commercial firms secure legal rights to inventions. In the United States, the core economic justification of patents is that they improve welfare by stimulating discovery, disclosure and dissemination. However, patent systems generate both economic benefits and costs. In principle, the theoretical effects of patents on innovation depend on the institutional environment and on the nature of technology. Both academics and policy makers are aware of circumstances in which patents can adversely affect innovation. For example, in some situations, the patent system has the potential to create a thicket of fragmented, overlapping property rights which raises the costs of innovation.

This chapter explores one alternative to the reliance on patents: 'free and open' development. We refer to a system that relies on free and open development as an 'open innovation' system. Open innovation systems side-step the patent thicket problem by not asserting patent rights to inventions. Open innovation systems rely on the free sharing and dissemination of knowledge. The idea that an open innovation system could successfully foster innovation may seem outlandish: where are the necessary incentives for innovation? We therefore offer a tentative discussion of how open innovation systems can work (that is, of the incentives and governance structures). We also discuss some of the advantages and disadvantages of each system in stimulating inventive activity.

As a concrete illustration of the issues, we consider the software industry. Substantial reliance on software patents is very new in this industry, and it is an industry where thicket problems seem likely (due to the nature of software inventions and patents). It is also an industry where free and open develop-

ment activity has been intense. Our examination of the software industry leads us to tentative generalizations: principles and lessons that could apply to other industries and technologies. We present these in the hope of stimulating further debate.

The chapter proceeds in six further sections as follows: we provide some brief background on the justifications of intellectual property rights; we provide some background on patent law and institutions; we discuss some of the economic aspects of patent protection; we explore the problem of patent thickets; we discuss the analytics of open innovation; and we offer some concluding remarks.

Background

From an economic perspective, property is essentially a collection of enforceable duties and privileges. We call these enforceable duties and privileges 'property rights'. Property rights are a fundamental determinant of economic activity. Institutions of exclusion, which need not be legislated or even explicit, bring into existence property and motivate the creation of property. Equivalently, these institutions of exclusion embody the current property rights that make it rational to create new property.

The economic perspective on property is close to the legal view but perhaps far from everyday parlance, where we call a piece of land, or a house, or a shirt 'property'. This common usage serves as convenient shorthand and often creates little confusion. For example, a couple who buy a residence are unlikely to be surprised that they have not acquired the right to build a toxic waste dump on it. On the other hand, they may be surprised to learn that they have also acquired a set of upkeep duties imposed by local legislation. This indicates the imprecision inherent in everyday parlance.

Confusion grounded in such imprecision may occasionally sidetrack discussions of public policy. Policy discussions of property rights often refer to legislating the protection of property or, in an even more misleading turn of phrase, the protection of property rights. From an economic perspective, such rhetoric obscures the most fundamental and challenging aspect of property rights: property is *created* by ever-changing human institutions that render feasible the maintenance of various restrictions on human behavior. Property does not exist prior to these institutions. The institutions generate the property rights and thereby the property. Public policy modifies property rights by modifying these institutions; it does not simply determine the 'protection' of pre-existing rights. Public policy in the area of property rights is a project of creation and destruction, not of discovery.

Economists are particularly interested in public policy that influences the distribution and creation of wealth. This chapter reflects that emphasis. Of course, the related question of how economic considerations influence the

structure of property rights is also very interesting. The enclosure movement in England reflected the economics of the woolen industry (Weber 1923 [1981]). Native American property rights institutions changed in response to the fur trade stimulated by European colonists (Demsetz 1967). Property rights changes in response to increasing urban density in nineteenth-century America include the loss of the long-established easements of light and air (Friedman 1985, p. 413). Even now the boundaries of patentability and of the concepts of fair use of copyrighted material are shifting in response to the rise of information technology in the late twentieth and early twenty-first centuries. Clearly, changing economic conditions often lead to the creation and reallocation of property rights.

This chapter speaks of property rights as social constructs. Others speak of property rights as fundamental constituents of a metaphysical moral reality. For example, many libertarians and classical liberals assert a 'Lockean' claim that there are natural rights to our persons and tangible property, and that the basic justification of governments is to protect those rights. Such views are important for public policy, whether or not they are correct (however that might be assessed), because they determine some of the political attitudes toward property rights and thereby influence institutions and legislation. For example, whether or not it is 'self-evident' that individuals 'are endowed by their Creator with certain inalienable Rights', the notion was certainly politically influential. Metaphysical, rights-based moral reasoning influences the structure of property rights – that is, the social structures of duties and privileges that constitute property.

Justifications of intellectual property rights

It is often said that IPRs 'protect creations of the mind', but what human creation is not a creation of the human mind? IPRs are embedded in institutions that restrict the use of inventions or creative works. The rhetoric of 'protection' appears to appropriate the ancient social legitimacy of certain kinds of property rights for a different application: the right to exclude behavior that does not directly affect one's ability to enjoy a good or service.

At least since the nineteenth-century patent debates, proponents of copyright and patent protections have often used the rhetoric of natural rights and talk about the prevention of theft.[1] The extreme version claims that the fundamental Lockean argument that a person naturally owns the fruit of their labor applies even if that fruit is an idea. This view would apparently imply a radical extension of IPRs far beyond the bounds of existing institutions. For example, it would apparently allow IPRs even for abstract ideas, even mathematical theorems, which current IPR regimes do not allow. Additionally, such metaphysical claims imply no bounds to the period of ownership, and

perpetual patents conflict with one of the primary pragmatic aims of the patent system.[2] While such pragmatic difficulties will certainly influence policy, they might appear irrelevant for metaphysical reasoning.

Even at the metaphysical level the argument appears problematic, however. Most obviously, it fails to explain why patent rights should be able to exclude independent development of an innovation. Furthermore, it is far from obvious that the usual Lockean reasoning can be extended to non-congestible intangibles such as intellectual property, especially when the assertion of such property rights diminishes the property rights in tangibles. The application to IPRs does not adequately ponder what 'fruits' are by right enjoyed by a creator: those of use, or those deriving from exclusion. These sets are peculiarly non-overlapping in the arena of intellectual property. Property rights in a tangible good include rights to restrict others' access because their access can interfere with the owner's enjoyment of the specific good. In contrast, IPRs are rights to restrict others' access to an instantiation of an idea, even when that access constitutes no interference with the owner's enjoyment of the creation. In the case of patents the attenuation of the rights of others is particularly obvious. The entire intent of patent protection is to restrict what others may do. Patent ownership does not even ensure the owner a right to use the patented technology – since its use may be dependent on other patented technology – but merely grants a right to take alleged infringers to court (Thomas 2002).

The pragmatic view stresses utilitarian rather than metaphysical considerations. The core pragmatic claim for IPRs is that innovation will be undersupplied in their absence. For example, Romer (1990) suggests that innovators must anticipate a period of monopoly rents to justify the sunk costs of innovation – an argument often attributed to Joseph Schumpeter. The core pragmatic justification of IPRs therefore claims the presence of a trade-off. By means of IPRs of limited duration, society provides innovators with potential rents in order to increase the production and diffusion of innovations. Supportive claims specific to patents are that IPRs increase the disclosure and dissemination of useful knowledge: strong patent protection may reduce the need for invention secrets. Patents are sometimes characterized as a social contract, wherein patent protection is granted in exchange for the surrender of invention-specific trade secrets.

Dynamic considerations are clearly fundamental to the pragmatic view since typically innovation takes place in one period and compensation in a future period. From a static perspective, the temptation is not to provide IPRs once an innovation is developed, for there are no incentive effects left – only rents. Since uses of an idea are non-rivalrous, the optimal policy appears to be to allow free use. Thus static considerations speak against rights to exclusion. From a dynamic (repeated game) perspective, this conclusion is flawed

because if innovators do not expect compensation, there may be no innovation and therefore no new ideas to freely distribute.

Intellectual property policies face a time-inconsistency problem: *ex ante* it is optimal for the authorities to promise exclusive rights (to stimulate innovation), but *ex post* it is optimal for them to deny exclusive rights (to allow free use). Brute promises are not credible: unless policy makers are institutionally constrained, innovators will disbelieve promises of legal protection. Pragmatically speaking, this suggests that policy makers cannot optimally stimulate innovation while systematically revoking IPRs.

If reputation effects were absent, a state might best promote the material standard of living by revoking all existing patent protection and then promising never to behave that way again. Of course pragmatists find this observation irrelevant: a state that revoked IPRs once could not avoid reputation effects (namely, the expectation that it would so act again). Advocates of metaphysical property rights find such proposals not just irrelevant but immoral.

IPRs are sometimes characterized as a pragmatic response to the failure of free markets to supply adequate innovation, but IPRs like other rights are simply one of the conditions that determine the nature of markets. Like property rights in tangible goods, IPRs create markets. Economists are inclined to wonder whether alternative market solutions would fail to emerge in the absence of IPRs. For example, what might prevent innovators from contracting with the potential beneficiaries of the innovation? The most obvious answers concern transactions costs and moral hazard problems, so these appear to be good areas to focus any policy discussion of IPRs.

Metaphysical and pragmatic justifications of IPRs both rely on presuppositions. Natural law justifications of IPRs rely on non-empirical claims about the essential nature of property rights. Pragmatic justifications rely on empirical claims about the net effects of IPRs on creative activity. It is important to the pragmatic justification that IPRs not only yield creative activity that would otherwise be inefficiently quiescent but that IPRs do so in sufficient quantity to justify the costs introduced by the system. For example, pragmatic justifications of patent systems assume positive effects on the discovery, disclosure and dissemination of useful knowledge that exceed the costs of the patent system in terms of the deadweight losses (deviations from marginal cost pricing) and resource costs (of operating the system).[3] Empirical assessment of this assumption is a challenge, largely because of measurement issues. There is much innovation that occurs without patent rights (particularly for innovations that are not easily imitated) and empirical analysis needs to control for this.[4] Furthermore, welfare assessments need to control for the fact that the incentives for private generation of information may in some cases be excessive, so that the patent process induces costly, duplicative resource expenditures (Hirshleifer 1971). For example, firms who could work

together to more cheaply produce a non-rivalrous input might instead waste resources in a patent (winner-take-all) race.

In the United States, the legal justification for IPRs is closely related to the pragmatic economic justifications. US patent law is based in Article 1 section 8 of the US Constitution, which gives Congress the power 'To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries'. This is therefore an important criterion by which to judge the efficacy of modern intellectual property institutions.

Contemporary patent laws

This section provides a brief introduction to patent laws. We cover the basic issues of patentability and the conditions for patent grants. Two key points are central to our discussion. First, patents are intended to protect not ideas *per se* but rather innovative practical applications of ideas. Second, patents are intended to promote access to ideas and to the know-how behind the application of ideas.

In most countries, the government makes the laws governing patents, the courts interpret them, and a government patent office implements them. The patent office receives applications for patent protection and decides on whether to grant a patent. In what follows, we speak primarily to US patent law and mention practices in other countries where relevant.

What is patentable?

The distinction between ideas and applications of ideas lies at the heart of patent statutes. Patents can be awarded for new solutions to specific practical problems. Products and processes are patentable. Abstract ideas or mere suggestions are not patentable. For example, $E = MC^2$ represents an unpatentable abstract idea, but a heat conduction system that applies this principle of atomic energy is patentable. Two or more inventions can be patented that use the same idea, so long as the inventions are sufficiently distinct as to avoid infringing each other (or contain infringing material). More on this later.[5]

These considerations are evident in Section 101, Title 35 of the US Code (Patent Act): 'Whoever invents or discovers any new and useful *process, machine, manufacture, or composition of matter*, or any new and useful improvement thereof, may obtain a patent therefore, subject to the conditions and requirements of this title' (emphasis added).

In practice, distinguishing between ideas on the one hand and products and processes on the other can be a challenge, particularly in some new technological fields (such as software and biotechnology). Patent offices and the courts nonetheless must make this distinction in their decisions or rulings. Nobel prizes, academic tenure, or other rewards may provide incentives for

abstract research, but the patent system does not (and is not intended to) play that role. This is reflected in international laws, which list the following exclusions from patenting (WIPO 1998): discoveries of materials or substances already existing in nature; scientific theories or mathematical methods; plant or animal varieties, or essentially biological processes for the production of such plant or animal varieties, other than microbiological processes; schemes, rules or methods, such as those for doing business, performing purely mental acts or playing games; and methods of treatment for humans or animals, or diagnostic methods practices on humans or animals (but not products for use in such methods).

What can be granted?

The above discussion focused on subject matter that is patentable. For a patent to be granted, however, the patent applicant should demonstrate the novelty, utility, and non-obviousness of the innovation.

An invention is not novel if it has been known, practiced, used, or sold previously anywhere in the world. For example, novelty is destroyed if the inventive idea has been published in a journal, book or dissertation, or practiced by a firm, cooperative, public enterprise, or individual. Any invention that is already in the public domain cannot be patented. An invention is also not novel if a patent application for the same (or similar kind of) invention was accepted or is pending elsewhere in the world. Inventors can also forfeit the right to a patent if they display the invention in a public place (other than for the purpose of testing a prototype at a well-recognized exhibition).[6]

An invention has utility if it is capable of industrial application. This reflects the point made earlier that patent protection should not be available for purely abstract ideas or creations. Patents on genetic discoveries are particularly controversial because in many cases it is not known what functions certain genes play or what utility they provide.

The non-obviousness of an invention pertains to the inventive step (or 'quality jump') of an invention; that is, its inventive contribution (or value added) to existing inventions or technical knowledge. For example, Section 103 of the US Code does not permit a patent if 'the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains'.

Obvious inventions are not only not novel, they may infringe on existing patent rights. The inventive jump must be large enough not to infringe existing patent rights. The required distance from existing patents – often called patent 'scope' – varies substantially across jurisdictions. In the United States, patent scope is broad. In Japan, patent scope tends to be narrow: patenting of

minor inventions around an existing patent is permitted. Patent holders enjoy greater market power in a system where patent scope is broad. Entrants find it easy to enter markets where patent scope is narrow.

Patent scope also helps determine the extent to which two or more inventions can build on the same idea (or on each other). In the United States, a litmus test is provided by the 'doctrine of equivalents', which holds that inventions that substantially perform the same function, in substantially the same way, and produce substantially the same result, are the same inventions. The doctrine allows the claims of an invention to cover not only those things that are *explicitly* stated in a patent document but also those that are *implicit*. Under the doctrine, an accused is liable for infringement for using the essence of a patented invention without literally infringing it. As a weapon of litigation, the doctrine of equivalents can be exercised by patent holders to make technologically neighboring inventions liable for infringement. This pressures rival firms to distance themselves – in 'technological' space – from the rights holder.

Enablement requirements of a patent

Each patent application must usefully disclose the details of the invention. The standard of disclosure is that a person skilled in the art can replicate and use the invention (without undue experimentation). This is the 'enablement' requirement. The inventor is required to explain not the best possible implementation but rather the best that he is aware of at the time of applying for a patent. Lack of fulfillment of these requirements is grounds for patent rejection (before the fact) or patent invalidity (after the fact). In the United States, the standard of information disclosure required by law is provided by Section 113 of the US Code (and is representative of other national laws):

> The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to *enable* any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the *best mode* contemplated by the inventor of carrying out his invention. (Emphasis added)

Patent applications are published 18 months after the date of application, whether or not patent rights are subsequently granted. If a patent is not granted for reasons other than infringement, the invention immediately enters the public domain. If a patent is granted, the idea immediately enters the public domain, and the invention enters the public domain when the patent expires.

From the social contract perspective, knowledge dissemination is a crucial quid pro quo for patent protection. By implication, the enablement requirement is a crucial component of the patent system.

Limitations and other conditions

Policy authorities may impose other conditions or limitations on patent rights. Examples include working requirements and compulsory licensing.

In contrast with the current US system, some jurisdictions have working requirements. Failure to 'work' the invention (by manufacturing or marketing it) within a specified time period may lead the patent holder to forfeit his rights to the invention or face compulsory licensing (that is, a mandate that the patent holder license his invention to third parties). While working requirements have been justified as forbidding the hoarding of technology, they clearly reduce the strength of patent rights by reducing the option value of patents. It restricts the right of the patent holder to choose the most opportune moment to market or manufacture his invention. For example, a patent holder might be temporarily financially unable to work the invention or might find current market conditions to be unfavorable.

Compulsory licensing may also be imposed in response to antitrust or anticompetitive actions of patent holders. This power may be used to limit patent blocking. Blocking arises when different patents cover subject matter in a way that manufacturing one good would cause an infringement of the patent rights of others. Blocking may arise if one technology is an improvement over another and where the improvement cannot be practiced without the use of the original, core technology. If the rights to the improvement and to the original technologies are owned by different patent holders, the original owner could block the improver by refusing a license or demanding 'unreasonable' terms. When the parties cannot resolve the problem through private negotiation, the parties can turn to the court system. Conflicts such as these take matters out of the sphere of intellectual property law and into that of competition law (for example, the US Sherman Act, Section 2).

Another anticompetitive situation that might arise is if one patent holder owns the rights to a technology that serves as an 'essential facility' for other producers in the market or downstream. In certain environments a specific computer operating system may prove to be an essential facility.[7] A similar situation arises if a technology owned by a patent holder becomes a *de facto* standard for an entire industry. In these cases, patent rights extend the market power of the patent holder considerably. Again, these matters come under the purview of competition policy.

Patents create a temporary right to exclude others from practicing an invention. While this may in some situations result in a monopoly producer or supplier of a good or input, there is nothing illegal about being a monopoly. Rather it is illegal to engage in conduct that maintains and extends market power. Determining when antitrust considerations should supercede IPRs is a difficult matter. Situations where patent holders deny access to an essential facility, block the exploitation of other technologies, or refuse to

license or to deal, are cases where patent rights might come under antitrust scrutiny. In these instances, competition policies could limit the exercise of patent rights.[8]

Application: software patents

The general considerations raised above find an interesting application in the area of software-related patents. In the closing decades of the twentieth century, software production became a substantial source of economic value. In 1972 the Supreme Court ruled that mathematical algorithms are non-statutory subject matter.[9] The US Patent and Trademark Office (PTO) consequently considered computer programs to be unpatentable, like laws of nature or mathematical relationships. As a result, copyright and trade secrecy were the primary tools of software-related intellectual property protection.

This situation changed radically in the 1980s. In 1980 the Supreme Court asserted that Congress intended 'anything under the sun that is made by man' to be patentable subject matter. In 1981, the US Supreme Court asserted essentially that, although algorithms are not patentable, software that has a technical effect is patentable.[10] Critics consider this an odd ruling: the court appears to rule that combining admittedly non-statutory subject matter (the software algorithm) with known art (the curing frame) yielded a patentable 'process'.

The struggle to clarify the patentability of software led in 1996 to the PTO's *Examination Guidelines for Computer-related Inventions*, which made the possession of a 'practical application' the key criterion for patentability of the software:

> A process that merely manipulates an abstract idea or performs a purely mathematical algorithm is non-statutory despite the fact that it might inherently have some usefulness. For such subject matter to be statutory, the claimed process must be limited to a practical application of the abstract idea or mathematical algorithm in the technological arts. For example, a computer process that simply calculates a mathematical algorithm that models noise is non-statutory. However, a claimed process for digitally filtering noise employing the mathematical algorithm is statutory. (PTO 1996, Section IVB.2.b.ii)

The result has been an increase in software patent applications as well as increased criticism of the appropriateness of some of the patents granted.

Critics believe that the courts and the PTO have tried to draw a line where none exists, since software applications are inherently a sequence of mathematical computations. The PTO and the courts, in turn, have insisted that there is a fundamental distinction between mathematical algorithms and other software-driven computer processes.

Prominent computer scientists (notably Donald Knuth and Richard Stallman) reject this view and have asserted that patents are being granted for program-

ming techniques likely to be found in the homework of first-year computer science courses. They suggest that patent examiners fail to see the obviousness of many software-related inventions, and have overlooked much prior art in the field of software innovation. As Stallman (2000) has emphasized, there is also a sense that the PTO's limitations are being deliberately exploited by patent applicants who wrap simple and obvious ideas in complicated phrasing to make them look like patentable inventions.

A converse source of controversy concerns the relationship between software patents and the enablement requirement. As discussed above, disclosure is a quid pro quo demanded by society for the benefits provided to the patent holder. As in other industries, the availability of software-related patents is supposed to encourage inventors to disclose the underlying technical details of their inventions. Unfortunately, it remains an open question whether the intent of the disclosure requirements is fulfilled in the case of software-related patents. First, like other patents, software-related patents grant 20 years of patent protection. However, in such a rapidly developing field, many observers have argued that a shorter patent duration or a narrower patent scope may be more appropriate. Second, the disclosure of a software program's source code has not been required for satisfaction of the enablement plus best mode requirement. Indeed, applicants are 'encouraged to functionally define the steps the computer will perform *rather* than simply reciting source or object code instructions' (emphasis added). A more obvious disclosure requirement would be for a functional definition, ideally in a specified and widely used modeling language (such as the Unified Modeling Language), in *addition* to source code provision.

It may seem that the lack of a source code disclosure requirement results in an inadequate disclosure standard for software patenting.[11] However, one may argue that source code disclosure eliminates the reasonable experimentation that may be required to implement an adequately disclosed invention, so the requirement of source code disclosure would create an excessively demanding standard of disclosure for software inventions. Court rulings indicate that the enablement requirement of software patents varies by the nature of the invention and by the type of computer program needed to carry it out. If it is relatively straightforward for a programmer skilled in the art to write a computer program to carry out the invention, then the source code of the patented invention is not central to enablement. The duty of the patent holder in such cases is to be very clear, transparent and specific about what those steps are.[12] Indeed, the court rulings suggest that disclosure of the functions of the software is generally adequate because, 'normally, writing code for such software is within the skill of the art, not requiring undue experimentation, once its functions have been disclosed' (quoted from Lemley et al. 2003, p. 210). In contrast, if undue experimentation (that is, a lot of trial and error)

on the part of a skilled computer programmer is required before the programmer can repeat the invention, the source code must be disclosed. For instance, in one case (*White Consolidated Industries v. Vega Servo-Control*), the court ruled that enablement was unsatisfactory because it would take a skilled programmer about two years to implement the invention described in the patent specification.

These court rulings raise a very interesting question in cases where source code is not judged as crucial to enablement. If writing the necessary code is 'normally' within the skill of the art, can the software invention still pass the test of non-obviousness? This suggests that there should be a substantial burden of proof for any patent applicant who claims non-obviousness for a software-related invention but simultaneously insists that source code provision is unnecessary. The ability of a programmer to do without the source code suggests that any skilled programmer approaching the same problem might produce the same solution: that is, it suggests that the invention is obvious to someone with normal skill in the art. Thus software patents appear to create a quandary: either the system is permitting the patents of some obvious things or it is relaxing the enablement requirement.

Some economic aspects of patents

The previous section focused on a few legal and institutional considerations of patents. We now turn to economic considerations. The theoretical literature on the economics of patents is rather large and varied, and we consider only a few salient economic considerations.

Why firms patent

The decision to patent depends on economic factors, such as market conditions, government policy, costs of patenting, and reputational, signaling and other strategic considerations. An inventor will seek patent protection for an invention if the net benefit of procuring patent protection exceeds the cost of filing for protection. We may represent this condition as $v_p - v_o \geq c$ where v_p and v_o denote the value of an innovation with and without patent protection, and c denotes the total cost associated with filing a patent (including, for example, the present value of any maintenance fees). The value v_o captures the best alternative to patent protection, including any steps taken to appropriate the rewards from the innovation (such as lead time, reputation and secrecy). Thus the value of a patent right is the incremental return $(v_p - v_o)$. A patent application is filed when this value exceeds the (total) filing costs.

The conventional wisdom is that firms demand patent protection in order to safeguard their intangible assets, which imitators might copy and distribute at nearly zero marginal cost without incurring any sunk development costs. As we have discussed, a standard justification of patents is that they may

secure gains to innovators and thereby ensure adequate incentives to innovate.

Recent survey results challenge the conventional wisdom. For example, Levin et al. (1987) report that firms do not, in general, regard patent protection as very important to protecting their competitive advantage (and thus to appropriating the returns to their investments). Alternatives such as trade secrecy, lead time, reputation, sales and service effort, and moving quickly along the learning curve are more important.

If patent protection is not important as an instrument for appropriating the returns to innovation, why do firms patent (and patent a lot)? Cohen et al. (2000) report a variety of reasons why firms patent: to block rivals from patenting related inventions, to gain bargaining chips in cross-licensing agreements, and to measure internal performance (of the firms' scientists and engineers). Their survey indicates that these factors are more important determinants of patenting than the direct protection of research and development (R&D) investment returns.

These survey findings should be viewed as provocative rather than definitive. First, the importance of patents (and purpose of patenting) varies by industry, being very important for chemicals and drugs. Second, the responses of firms (or their attitudes towards patents) may have been influenced by the patent regime in place: it may be easier to dismiss the importance of patent protection when it is readily obtained. Third, the responses of interviewees may not be fully comparable: the interpersonal meaning of numerical rankings is uncertain. This makes it difficult to tell whether the responses reflect differences in firm behavior or random errors. Fourth, the surveys are based solely on US firms' experiences. A similar comprehensive study for Europe and Asia would shed additional light on patenting behavior. For example, the strength of patent regimes varies much more internationally than domestically, so international data may be able to address our second concern above. Despite the methodological and theoretical problems, the available survey evidence is provocative and should stimulate further inquiry about the patent system.

A simple, illustrative model

We shall model one way that patents can be productive or counterproductive, depending on the nature of technology. We choose a model that makes no attempt to introduce dynamic or strategic considerations. It is an essentially static 'one shot game' model adapted from Bessen and Maskin (2000).

Consider a single firm that can engage in R&D. By sinking cost c_1 the firm has a probability of research success p_1. The expected value of undertaking the research is $p_1 v_1 - c_1$, which is of interest only in the case where this is positive. Following Bessen and Maskin, we keep things simple by assuming

that this firm can capture the social value of its innovation, so that R&D takes place if and only if it is socially desirable.

Now introduce a second firm, a pure imitator that produces an imperfect substitute for the first firm's product. We allow imitation to be costless to the imitator, and after imitation the innovating firm only captures a fraction $s < 1$ of the social value. The net payoff to R&D falls to $s\ p_1v_1 - c_1$. If this is negative, socially desirable innovation does not take place. As Bessen and Maskin note, this is the classic case for patents. Without patents, it may not pay a firm to invest even if it is socially desirable. Patent protection can ensure the ability of the first firm to capture the social value of its innovation, thereby ensuring that a socially desirable research expenditure takes place.

Coase (1960) is often cited in support of the proposal, often called the 'Coase theorem', that in the absence of transactions costs the allocation of property rights does not affect efficiency. This suggests that efficiency-based arguments for IPRs should be explicit about how the Coase theorem fails to apply. For example, in our model of the classic case for patents, an alternative to patent protection might be for the innovating firm to pay the imitating firm not to imitate. This is another arena where intellectual property law and antitrust law interact: such a negotiation may not be allowed by antitrust law.

Now suppose $s\ p_1v_1 - c_1 > 0$. In this case patents are not needed to induce R&D, since the firm can recoup its investment without patent rights. In this simple model, the pragmatic justification for patent protection disappears, but patent protection is neither harmful nor productive from a social welfare point of view. (In a dynamic context, however, the level of R&D investment is an endogenous decision. The expected level of profits should influence how much R&D the firm chooses to undertake.)

Bessen and Maskin (2000) note that patent protection may still prove beneficial if firm 2 is not a pure imitator but instead is symmetric in its ability to produce R&D. By assuring economic rents to one firm, patent protection can induce both firms to invest in R&D.

Let us consider a different situation. Assume $s\ p_1v_1 - c_1 > 0$ so that firm 1 will conduct R&D even in the absence of patent protection. But this time suppose firm 2 is also an innovator that is working on a complementary innovation. That is, the innovation has value only given access to the innovation of firm 1, so that a patent granted to firm 1 will grant a holdup right over firm 2. Firm 2 can sink cost c_2 to produce a probability p_2 of research success. The innovation has a social value v_2, which firm 2 can capture. If firm 1 does not have a patent right, firm 2's expected value of undertaking the research is $p_1p_2v_2 - c_2$. We consider the case where this is positive, so that research is worthwhile to firm 2.

If firm 1 has a patent right, then subject to the constraint that firm 1 will not lose by licensing its technology, firm 2's expected value of undertaking the

research is no more than $p_1p_2[v_2 - (1 - s)v_1 - t_2] - c_2$, where firm 2 now pays a license fee to firm 1 (at least $(1 - s)v_1$) in order to produce and where t_2 is the transaction cost of negotiating a patent license. The new expected value of undertaking the research may be negative if the license fee is too high, the innovation value too low, or the transaction costs too high. The problem arises when that firm 2 cannot afford to compensate firm 1 for the decline in firm 1's ability to capture the social value of its innovation. In this case, patent protection allows firm 1 to block firm 2's innovation. Firm 2's valuable innovation is subject to 'holdup' by firm 1.

Patent thickets

Shapiro (2000) defines a 'patent thicket' as 'an overlapping set of patent rights requiring that those seeking to commercialize new technology obtain licenses from multiple patentees'. For example, he notes that a semiconductor manufacturer can potentially infringe on hundreds of patents with a single product. In such circumstances the transactions costs of negotiating licenses with many different patent holders might prove prohibitive even for a valuable commercial innovation. Shapiro argues that the perceived danger of holdup has introduced a threat to innovation that is 'of first-order significance'.

The size of the transaction costs that must be incurred to negotiate a complete set of licenses is one possible measure of the density of the thicket. These costs will vary with industry structure and also with the nature of the innovation process within an industry. However, even aside from the transaction costs dimension, patent thickets are likely to be socially costly. Antoine Cournot demonstrated that the behavior of multiple input-supplying monopolists can lower total profits and simultaneously reduce consumer welfare. Shapiro (2000) extends this analysis directly to patent thickets and argues that coordination among the input suppliers can therefore improve social welfare. Buchanan and Yoon (2000) generalize this result and show that multiple rights-to-exclude generally lead to resource underutilization, as suggested by the Heller (1998) discussion of the 'tragedy of the anti-commons'. Patent thickets introduce the possibility of a perverse outcome in the patent system: in a classic tragedy of the anti-commons, a system whose acknowledged purpose is to promote innovation may produce patent thickets that stifle it.

Where innovation tends to be highly incremental and cumulative, essential licensing for a new innovation is more likely to involve many patents. If only a few firms hold most of the fundamental patents in an industry, or if firms can form an adequate patent pool, the per patent transaction costs involved in licensing should be diminished. However, industries in which innovation is highly incremental and cumulative may prove particularly likely to see high dispersion of innovations across firms and even across private individuals.

(Software and biotechnology, for example, appear susceptible to such dispersion.) Such industries are therefore likely to develop dense patent thickets. In contemplating the possible costs of such patent thickets, one may analogically consider the field of mathematics: developments are highly incremental and cumulative, and few observers would propose burdening mathematicians with the requirement that they negotiate a license for each extant theorem used in a new proof. Are lessons from the vigorous field of mathematics, where patenting is forbidden, relevant to other areas of incremental and cumulative creative activity?

Cooperative arrangements can mitigate patent-thicket problems. For example, large manufacturers often cross-license their large present and future patent portfolios. This may seem a promising and potentially pro-competitive arrangement, but smaller and newer players may be stymied. Cross-licensing among larger corporations can promote collusion and the exclusion of competition from new entrants (Boldrin and Levine 2002). For example, semiconductor firms accumulate large patent portfolios that they then cross-license with rivals (Hall and Ziedonis 2001). This is clearly another area where patent policy and antitrust policy overlap: antitrust law historically has viewed with suspicion mechanisms for cooperation among competitors. This creates a policy dilemma: the cooperation necessary to avoid patent thickets may be precluded by antitrust concerns. Shapiro (2000) proposes that policy makers can ensure that cross-licensing is pro-competitive by attending to the nature of the patents involved: if patents licensed together are complements, not substitutes, then cross-licensing is more likely to be socially beneficial.

Patent pools offer another potential solution to patent thickets. Firms with interlocking patents may form an organization for the purpose of sharing patent rights. The organization could be an open pool, where members license patents to one another and to third parties, or a closed pool, where members license only to one another. An important superiority of patent pools over regular cross-licensing agreements among firms is that the pool acts as an entity *vis-à-vis* licensing to outsiders. In a cross-licensing agreement, one party has the right to use another's patent but not to sublicense it to a third party.

Once again antitrust considerations arise. As with any joint venture or cooperative arrangement, patent pools may generate opportunities to behave as a cartel or fix prices (for example, royalty rates on licenses). Yet patent pools may serve the public interest if they integrate complementary innovations, promote knowledge sharing, and reduce the transaction costs of negotiating licenses. Regulatory accommodation of patent pools has fluctuated over time, and current acceptability was probably influenced by the prevalence of complementary innovation in the computer industry (Lerner et al. 2003). Current law appears to reflect these issues of substitutability and

complementarity: US Antitrust Guidelines permit patent pools among 'essential patents' – that is, among patented inventions that are complementary.[13] Another requirement of these guidelines is that members of a pool retain the right to license independently, which helps destabilize any collusive behavior.

Standards

The emergence of formal or informal industry standards allows for coordination of industry development efforts. Wide adoption of a standard increases the value of patents essential to implementation of the standard. If unrestricted, a firm may realize this value directly through high licensing fees or indirectly by excluding competitors from implementing the standard.

Industry groups can form standards development organizations (SDOs), which strive to reconcile the interests of intellectual property owners with the interests of others who wish to practice the standard. One common way SDOs achieve this is to require participants to disclose their patent interests in the standard and to license all patents essential to compliance with the standard on 'fair, reasonable, and non-discriminatory' terms.[14]

Technology standards promote interoperability, even in the absence of direct contact between developers. This can reduce problems of 'lock-in'. Standards may also offer an escape from the patent thicket. Technology standards with publicly known licensing terms can reduce patent thicket problems. Unfortunately, as Shapiro (2000) stresses, concerns about antitrust actions have at times left the details of these terms to be determined *ex post*, after leverage is acquired by the incorporation of the patents into the standards. Standards offer a potential for escape from the patent thicket, but standards setting bodies generally need cooperation from the antitrust institutions: *ex ante* price limitations are crucial.

Free and open standards offer these gains on a non-discriminatory basis and also provide clear *ex ante* price limitations. Free and open standards are by definition available for all to read and implement without payment. The term 'free' refers to the implied freedoms, not the price. The term 'open' refers not only to the open accessibility to the standard but also to the openness of the process, which is intended to preclude the promotion of the market power of specific vendors or groups.[15]

The benefits of open standards to industry participants, especially new entrants, are clear, but economic analyses of the incentives to actively participate in open standards bodies remain incomplete. Economists need a fuller understanding of the incentives of intellectual property holders to initiate and participate in the development of open standards, which appears to involve the private provision of a public good. If a given industry is developing open standards, a firm may of course choose active participation as a way to ensure alignment of its development process with emerging standards. This suggests

that open standards movements become industry dominant if they gain 'critical mass'. In any case, it is clear that open standards at times garner widespread industry support.

A well-known example of a successful open standards body is the World Wide Web Consortium (W3C), the preeminent open standards body for the Internet. Hundreds of organizations participate in the development of interoperable technologies (specifications, guidelines, software, and tools). For example, the W3C's HTML (hypertext markup language) and CSS (cascading style sheets) standards affect anyone who 'surfs' the web. The W3C is also notable for its efforts to remain independent of specific vendors, who may try to co-opt an open standard through 'embrace and extend' tactics. The W3C patent policy is that its specifications must be implementable on a royalty free basis. The specific mission of the W3C Patent Policy Working Group concerns 'the growing challenge that patent claims pose to the development of open standards for the Web'. This open standards body has so far had remarkable success despite apparent vendor efforts to co-opt the standards. The widespread adoption of W3C standards has ensured a remarkable level of interoperability on the Internet and has supported an explosion of competing, standards compliant technologies.

Free and open development

To economists, explaining free and open (FO) development is even more challenging than explaining open standards movements. Once again the apparently simple modifier 'free and open' proves quite complex, but we shall focus on a few key features of FO development. In FO development, enabling disclosure is readily available, and licensing to use, redistribute and modify the technology is provided gratis by the developer. (Distribution of modified technology may be restricted to ensure that the modification also remains free and open.) Technology placed in the public domain obviously satisfies these requirements.[16] However copyrighted or patented technology can also be free and open.

When FO development succeeds in producing rapid innovation, this challenges the presumptions of many intellectual property arguments. Rosenberg (1976) documents FO development in the machine tool industry, von Hippel (1988) in the scientific instrument industry and Allen (1983) in the iron industry. An interesting feature of these cases, which appears typical of FO development, is that users of technologies were actively involved in the innovation process.

The most famous FO development effort has taken place within the software industry. Software development is considered to be free and open only if the source code is readily available and freely redistributable.[17] Software produced by FO development is generally referred to as free and open source

software (FOSS). FOSS development does not commit anyone to distributing source code for free; this would be inappropriate given that all known distribution methods are resource using. It does, however, mean that there are no legal restrictions on the redistribution of the unmodified open source software to others, and in practice it has meant that FOSS software has generally been provided for download without charge.

Software could be open source but adhere to few standards, open or otherwise. However, most well-known open source software projects stress adherence to open standards. To the puzzlement of many economists, this combination of open standards and open source development has generated tremendous economic value. Particularly famous open source software applications include the Linux operating system, the Apache web server, the MySQL relational database, the sendmail mail transfer agent, the Mozilla web browser, and the interpreters for the Perl and Python programming languages. These applications are all in wide use and are considered highly competitive with commercial products. They demonstrate unambiguously that high-quality, commercially important, and very innovative development can take place in the apparent absence of revenue-generating patent rights.

We shall use the term 'the open source phenomenon' to refer loosely to the success of FOSS development in producing economic value despite stringent limits on the ability of individuals and firms to appropriate the value that is being created. Economists recognize that explaining the open source phenomenon is an important project. We consider some explanations in the next section. Of immediate interest is that FO development offers clear potential for escape from the patent thicket.

A necessary condition for software to be considered 'open source' is that users and developers have the right to modify the code for their own use. Firms can therefore shield themselves from patent surprises by relying internally on proprietary or open modifications of free and open software.[18] To the extent that it can remain self-contained, FOSS development clearly eliminates the patent thicket problem.

While modified FOSS applications can be 'freely' distributed – that is, without a royalty or fee paid to the innovators of the original software program – these distributions are generally subject to one or more licensing agreements. A key aspect of these licensing agreements is the extent to which modifications of the code can be made proprietary. For example, the Berkeley Software Distribution (BSD) license allows the distribution of proprietary and closed modifications. The licensee is permitted to sell a modified program for profit (without having to reveal the source code of the modification). In contrast, under a General Public License (GPL), no proprietary rights can be asserted: distributed modifications must remain FOSS under the GPL.[19] In the early 1990s, the BSD license was the dominant form; at present, it is the

GPL. Lerner and Tirole (2002) indicate that use of the GPL appears to be declining in favor of licenses that permit proprietary modifications.

FOSS potentially offers firms an escape from the patent thicket. If the software is in the public domain, the firm can simply consider whether it pays to license patented technology given the FO alternative. The decision is similar with 'liberal' licenses such as the BSD license. However, the use of GPL'd software raises serious issues for the firm: it must consider whether this will require donation of innovations to the FOSS community beyond what will be repaid by the added insurance against stumbling into a patent thicket along with the ability to modify the GPL'd software for its own purposes.

Economics of free and open development

The open source phenomenon poses a puzzle. At first glance, it seems that economic reward has not been necessary to draw forth the factor inputs devoted to free and open software creation. Substantial economic value is being created despite the apparent absence of a price mechanism to direct this creation. Innovation appears to be substantial despite the apparent preclusion of appropriation of the returns to innovation.

Naturally this has drawn the attention of economists, who are particularly inclined to see prices as a necessary conduit of the information that can allow efficient resource allocation. Prices are a mechanism markets use to allocate resources among competing uses. Underpinning this is the role of property rights, which allows unambiguous reallocation of resources, goods and services. Together, prices and property rights provide incentives to manage resources in ways that produce economic value. Free and open development efforts appear at first glance to dispense with prices, markets and property rights in the process of value creation.

The present section addresses two issues. First, we consider some economic incentives that help explain and sustain FO development. In particular, we ask how FOSS is able to produce substantial economic value, apparently without the guidance of prices and intellectual property rights. Second, we ask whether FO development and strong IPR institutions (particularly patent rights) are mutually incompatible. In particular, we ask under what conditions patents adversely affect FOSS activity and under what conditions they might be complementary. While the focus of this discussion is the software industry, we attempt to draw some general lessons for FO development.

The FOSS Community

Economists are still struggling to understand the FOSS community. It is not at all clear that a single model of FO development will prove relevant to this varied community, which comprises private individuals, standards organiza-

tions, and firms of diverse sizes – all with varying degrees of involvement in FOSS development. In addition, governance structures vary from community to community. Some FOSS projects are run by strong, centralized leaders (as in the Linux project), while others are run by committees (as in Apache). Nevertheless some generalizations seem possible.

As Lerner and Tirole (2002) emphasize, FOSS innovations have not been just in the products themselves but also in the development process. FOSS development tends to take place in a collaborative organizational structure, and technically sophisticated users often provide important impetus for incremental FOSS innovation.

FOSS innovation is – and has largely been – a private sector initiative. Commercial firms – both start-ups like Red Hat Linux and behemoths like IBM – have been involved at many different levels. Private agents produce innovations using a mixture of private inputs (such as programmers' labor) and public goods (such as the public domain stock of technical knowledge). The FOSS projects are not state owned or conducted by state-owned enterprises, although they occasionally receive modest state subsidy.[20] Since FOSS products are understood to produce positive externalities, such subsidies might be amenable to traditional economic justifications. The depth of this justification is complicated when FOSS development directly competes with profit-oriented development, and it remains an active debate among policy makers and economists.

FOSS incentives

In FO development, it often seems that highly experienced labor works for free in order to provide sophisticated technical innovations at no charge. Economists are naturally interested in understanding why resources are being provided without compensation. Various proposals have been offered. Altruism is often mentioned both within the FOSS community and by observers, but it is a motivation that does not sit well with most economists. Economists generally seek more traditional motives: direct or indirect economic advantages gained by participation in FO development.

One proposal focuses on the role of users in improving extant technology. Incremental innovations that can be made at low cost by the user may have a substantial higher private valuation. For example, tinkering may pay off directly in the usability of a software application that has already been adopted. Of course, unless it is a GPL'd application, this does not explain why the innovator would eschew the assertion of property rights in the innovation. However, if one programmer's contribution leads others to invest in the project, his private valuation could increase (due to network effects) or costs decrease (due to a productivity effect owing to a higher stock of solved problems). This suggests that multiple equilibria are possible, depending on

the conjectural variations held by individual programmers. Game-theoretic aspects are thus potentially important in understanding FOSS incentives.[21]

A different explanation is offered by Lerner and Tirole (2002), who focus on career signaling and peer recognition effects. A programmer's contribution could enhance his reputation or peer recognition and thus lead to better future job offers or to venture capital. This suggests that programmers have a 'signaling' incentive to contribute to an open source project. It may be easier to signal in an open source project than in a proprietary project because the open source projects offer greater visibility.

Oddly, Lerner and Tirole discount the usefulness of altruistic and intrinsic motivations while citing an interview with Linux developer Linus Torvalds that highlights these (Torvalds 1998). Torvalds in contrast discounts career signaling and peer recognition effects, saying 'it feels good to have done something that other people enjoy using'. Torvalds even addresses compensation directly: 'I want to continue to try to avoid making money directly off Linux – that keeps me focused on purely technical issues with the Linux kernel'. This interview is remarkable because the interviewer pushes hard to pinpoint incentives to produce free software that an economist can fathom, yet he ends up with responses like: 'I really don't think you need all that much *quid pro quo* in programming – most of the good programmers do programming not because they expect to get paid or get adulation by the public, but because it is fun to program' and 'The first consideration for anybody should really be whether you'd like to do it even if you got nothing at all back'. A tendency to discount such altruistic and intrinsic motivations is common in the economics literature. Exceptions include Osterloh et al. (2003).

An interesting recent development is the contribution of large, established corporations to open source products. For example, the Open Source Development Lab has investment backing from numerous large firms, including Computer Associates, Fujitsu, Hitachi, HP, IBM, Intel and NEC. IBM in particular has been strongly backing Linux as a venue for selling services, applications and hardware. While important individual participants in FOSS development do seem to contribute in response to altruistic and intrinsic motivations, other motivations will be needed to explain the involvement of commercial firms.

If a user community is attracted to open source software, possibly for reasons of security or perceived quality, then FOSS development may still provide commercial interests with lucrative soil in which to grow. For example, it may be possible to profit by providing complementary products and services including accessories (such as computer manuals or utility applications). The strategy of 'giving away the razor in order to sell blades' is a common business practice in other industries, and it may go a long way in explaining the contributions of commercial firms to FOSS development. Per-

haps the best-known example of this is Red Hat, which charges for technical support on Linux-based products.

There can also be strategic considerations. Firms trying to free themselves from Microsoft's near monopoly on desktop operating systems may see a strategic advantage in promoting Linux development. Some governments as well have clearly acted out of such strategic considerations. Another possibility is that improvements in an FO product might enhance the market position of a related proprietary product. For example, Mustonen (2002) develops a model in which a monopoly producer of a proprietary software application supports GPL programs in order to achieve compatibility between its proprietary product and products that it cannot produce by itself. The monopolist gains from this strategy if sufficient network effects exist in the consumer market, so that the profits from compatibility exceed the loss in market share. Another possible motive commercial firms have to support FO development is to hurt firms that produce competing products. In contrast with predatory pricing, however, contributions to FO development cannot be repriced after successful predation. Even when there are no direct gains from freely revealing innovations, contributors to FO development may benefit if by revealing information they induce the development of desired complements to their proprietary production activities (Harhoff et al. 2002).[22] There are such a variety of reasons why the disclosure and free distribution of innovations might economically profit workers and firms that the research task of determining the most economically important is sure to prove challenging.

FO development versus IPRs?

The next set of issues concerns the relationship between IPRs, particularly patent rights, and FO development. We first discuss arguments for and against the proposition that patent rights threaten the open source movement. We then discuss how open source communities may employ the existing system of IPRs to protect their system of innovation.

Patent rights are a threat If patent rights increase the opportunity cost of participating in FO development, innovation may slow in the FO sector. If a rise in the strength of patent protection increases the returns to proprietary innovations, resources should shift away from the open source sector to the proprietary sector. Another concern is that a (non-GPL'd) project could be 'hijacked' if it were diverted into a proprietary commercial venture. Some people feel that this is what happened when AT&T began to assert copyrights over the use and distribution of Unix, after years of development in cooperation with universities and other research organizations. Another famous but less dramatic example occurred when the X Consortium, after years of en-

couraging volunteers making software submissions to reject the GPL, actively considered ending the free software status of the X Window System.[23]

A more substantial way in which patenting activities can harm open source activities is by fencing off certain software knowledge capital, preventing the exploitation of useful programming innovations. In addition, because the disclosure standards for software patents may not be fully enabling, access to the blueprint underlying the technical effects of software programs may be obstructed. Some firms for strategic purposes may hoard patents, never commercializing them, so as to prevent rivals from marketing close substitutes. Thus, open source projects may themselves be blocked by patent rights. The proprietors of patented technologies might demand royalties or fees that are too high (relative to the private benefits of the project) or refuse licenses altogether for strategic reasons.

A self-contained FO development project may offer a refuge from patent thickets and patent blocking. However, FO projects may find themselves unable to operate in a self-contained fashion. An FO development project may therefore find itself undermined by patent thickets. Just as with traditional development models, technology essential to a project may have been patented, and this may not be discovered until the FO development has consumed considerable resources. In sectors inclined to patent thickets, FO development faces difficulties similar to those faced under ordinary proprietary development. This shows up in struggles to maintain the integrity of the FO development process. That even FO development may face patent thicket problems may provide further support for the claim that patents are 'not appropriate for industries … in which innovations occur rapidly, can be made without a substantial capital investment, and tend to be creative combinations of previously-known techniques', as Oracle Corporation put it in a famous 1994 statement to the PTO.

Patent rights are not a threat The view that patent rights can readily coexist with FO development rests primarily on the case that the two approaches (proprietary versus free and open) have technologically separated niche markets. A related argument is that the two approaches specialize in different kinds of software technologies.

The 'niche' argument is that FOSS innovators develop customizable 'expert-friendly' products while proprietary innovators develop products for the mass market. There are several reasons for this. The programmer who participates in a FOSS project tends to be an expert at fixing bugs and adapting a program to suit his needs. The open disclosure of source code enables him to do that. The same programmer, as many have noted, has no marketing-oriented incentive to write user-friendly technical manuals. (Of course these could be written for profit under copyright protection.) Moreover, Bessen

(2001) argues that as software becomes more complex, debugging costs increase exponentially. Bessen argues that proprietary developers will therefore sacrifice complex, customizable features in favor of standardized products. Moreover, customizable products represent a small share of the market for any proprietary producer. Hence the open source community provides opportunities for particular consumers to customize the products themselves.

Evans and Reddy (2002) and Schmidt and Schnitzer (2003) discuss whether the lack of a profit motive lies behind the reason why FOSS developers do not produce goods for the mass market. The proprietary developer has incentives to undertake market research to figure out the needs of the consumers (particularly the least computer sophisticated consumers). Thus, they argue, proprietary producers better serve the end-user market, while the FOSS community better serves information technology professionals and other sophisticated users. We agree that the open source community has been extremely successful in producing expert-friendly software goods (such as operating systems, servers and programming tools) while proprietary producers have had greater success in producing user-friendly desktop applications. However, there are exceptions. FOSS development has turned more recently to the provision of user-friendly desktop applications, the best known of which is probably the OpenOffice office suite, and a variety of proprietary operating systems play important roles in many markets.

Intellectual property defenses The previous discussion focused on whether proprietary development under patent protection is likely to undermine or complement FO innovation. Patent rights appear less problematic for FO innovation when there is greater product differentiation between proprietary producers and FO developers or when FO developers work on innovations that are not patentable. When FO innovations are close substitutes for proprietary innovations, strengthened patentability appears likely to undermine FO development.

We now consider how FO communities use intellectual property laws to protect their system of innovation. For example, a natural defense against future patent thickets and blocking is to establish a robust set of documented prior art. This is widely recognized among FOSS developers, but there is disagreement about the best approaches to establish prior art. Some argue that invalid patents are being granted because of the inadequacy of the patent databases, which in turn is rooted in the extensive software development that took place before patentability, and that FOSS developers should therefore protect themselves by providing evidence of prior art to databases designed to facilitate patent search. Others argue that the opposite conclusion follows: invalid patents can be challenged, but making evidence of prior art more readily accessible also makes it easier to design a patent application that

avoids claiming prior art. Patent opponents therefore have no obvious incentive to provide such assistance. The more aggressive strategy of preemptive patenting suffers from the same weakness, although it provides stronger protection of specific technology. Instead, the argument goes, the FOSS community should provide *no* assistance to patent searches beyond the full and public documentation that ensure that prior art is demonstrable. This decreases the likelihood that software-related patents can be written in ways that avoid claiming prior art, and may therefore discourage or prevent firms from patenting. Unfortunately this defense strategy could result in socially wasteful litigation.

Preemptive patenting has additional drawbacks. It can impose substantial costs on the open source community – costs that cannot be offset by license revenues. Patenting is expensive in terms of the application fees, search and examination fees, attorney fees, translation fees if filed in or from other countries, and maintenance fees (if patent holders are required to pay fees during the life of a patent to keep the patent in force). Furthermore, the patents owned by open source communities may themselves contribute to patent thicket problems. For example, if a researcher wants to build upon a patented open source innovation and develop a proprietary product, the open source community may refuse a license or issue one on restrictive terms.

Finally FOSS developers do generally rely on non-patent IPRs, such as copyrights, to protect the source code of open source projects. They also use trademarks to protect the symbols and brand names of FOSS products. O'Mahony's (2003) survey of six major FOSS projects finds that the projects utilize copyright and trademark protection, software licensing, and other legal measures such as incorporation (in order to protect collective assets). The FOSS communities do not simply forfeit their intellectual property rights: they exercise them in a specific manner. These rights protect against misappropriation and help ensure that their innovations remain free and open.

Summary

Economic exploration of the functioning of free and open development has been stimulated by economists' recognition that FO development has generated substantial economic value in the form of technological innovation and diffusion. Economists' explanations have focused on subsuming the incentives for FO development in traditional economic categories. This valuable exercise has provided many insights into economic motivations for participating in FO development, but important aspects of FO development have yet to be given traditional economic explanations.

In FO development, innovations are fully disclosed and freely redistributable. It is clear that both pecuniary and non-pecuniary rewards can motivate participants of open source projects. However the pecuniary rewards are not tied to a

direct flow of licensing revenue. Innovators contribute resources to FO development without claiming revenue-generating IPRs. Modification of the innovations is unrestricted, although distribution of modifications may be restricted to keep them within the FO development community. FO development does not eschew use of the intellectual property institutions. Copyright laws, trademark rights and licensing help protect the creative assets and reputation of open source communities (against fraud, misrepresentation and 'hijacking').

The relationship between IPRs (particularly patent rights) and FO development is complex. It is conceivable that FO development could take place in a self-contained fashion, so that patent thickets elsewhere in the software industry would not hinder FO development. This is plausibly true in some niche markets. However, many FO and commercial development projects directly compete, and in these cases software patent thickets can pose quite a threat to FO development.

Conclusion

Intellectual property rights, like all property rights, are human creations. They are embodied in complex, interlocking institutions and sustained by pragmatic and metaphysical justifications. This chapter explores the IPRs embodied in patent systems. A core pragmatic justification of patents is that they foster the discovery, disclosure and dissemination of practical inventions. However, the patent system may in some cases produce complex overlapping property rights, particularly in sectors of the economy characterized by rapid, incremental and cumulative innovation. The resulting patent thickets can pose a costly barrier to the development of new technology.

Of course the phenomenon of holdup is not unique to intellectual property. Holdup also occurs in real property, as when a developer must purchase a large tract of land from different owners to produce an economic good but one of the owners holds out. Holdups are common not only when there are numerous parties involved but also when there are few parties but each has some market power.

Holdup is especially likely to be prevalent in the software industry. Software development is often a creative combination of known techniques, building on an extensive prior code base and working in concert with other programs or program components. If software patents render the code base, program components and programs all proprietary, programmers must obtain licenses from the owners of each component. This makes fertile ground for patent thickets. Moreover, software patents involve near-ideas: algorithms that produce technical effects are not always clearly separable from those that are simply mathematical algorithms. Because they protect near-ideas, software patents can potentially grant very broad powers, holding up follow-on inventions by restricting the use of near-fundamental discoveries.

There are several possible escapes from patent thickets, including cross-licensing and patent pools. Free and open standards and free and open development may also offer possible avenues for escape, and both have been used to create substantial economic value. Open innovation systems need not supplant but may operate alongside traditional patent systems. This would especially be the case for technologies that are not patentable in the legal sense (that is, in terms of novelty or non-obviousness), but there are situations where invalid patent rights (such as those for 'essential facilities' or for which prior art existed) could hold up FO development.

Despite a flurry of research activity in response to the startling successes of free and open source software, economic analyses of free and open standards and free and open development remain incomplete. Although no single explanation will encompass the diversity of free and open standards and development efforts, we suggest that free and open development holds particular promise for industries facing potentially severe holdup problems. Exploring the extent to which 'lessons' learned from free and open source software development can be applied to other industries should prove an exciting area for future research.

Notes

1. Machlup and Penrose (1950) note that many modern disputes were anticipated by the 'great patent controversy' of 1850–1875.
2. Machlup (1958) notes a 'permanent exclusive privilege' granted in Switzerland in 1577. Machlup and Penrose (1950) observe that nineteenth-century France and Belgium produced a substantial literature arguing for 'perpetual rights in intellectual products'.
3. As Machlup (1958) emphasized, consumers must pay more for any protected innovation, whether or not that innovation was patent-system induced.
4. Machlup (1958) proposes the early automobile industry as an example. We propose other examples in our later discussion of free and open development.
5. The Romer (1990) endogenous growth model correctly specifies the idea versus application of ideas dichotomy. In this model, the stock of knowledge is a public good while the intermediate capital goods whose designs or blueprints are derived from research knowledge can be made proprietary.
6. In some jurisdictions, such as the United States, the inventor is given a grace period of one year to apply for a patent after a public display.
7. In other environments local telephone networks, digital subscriber lines (DSL), or central railroad switching systems might prove to be essential facilities.
8. For a history of the relationship between these two types of laws in various countries, see OECD (1989).
9. In *Gottschalk v. Benson*, the US Supreme Court invalidated a patent on a computer program to convert signals from binary-coded decimal form to pure binary form.
10. In *Diamond v. Diehr*, the US Supreme Court upheld a patent on a process for molding uncured synthetic rubber into cured products. A computer algorithm was essential to the process, since it allowed precise and timely determination of the cure time based on the temperature of the molding press.
11. Of course, source code alone can also fail to constitute disclosure, since the source can be written to intentionally obfuscate the functionality. See Lemley et al. (2003) for further discussion.
12. For example, the *Examination Guidelines for Computer-related Inventions* of the PTO

suggest that enablement could be satisfied by the patent applicant by 'outlining the significant elements of the programmed computer using a functional block diagram'.

13. See US Department of Justice/Federal Trade Commission (1995), 'Antitrust Guidelines for the Licensing of Intellectual Property', http:www.usdoj.gov/atr/public/guidelines/ipguide.htm.
14. The Joint Electron Device Engineering Council (JEDEC) failed to get adequate disclosure when it admitted Rambus Inc. to membership in 1992, and Rambus apparently modified a pending patent to encompass the standards being developed by the JEDEC.
15. A more detailed introduction can be found at http://perens.com/OpenStandards/Definition.html.
16. Inventors can always put an invention in the public domain by publicizing it while refusing to patent it, as Jonas Salk famously did with the polio vaccine.
17. The source code for a software application is human readable (with a text editor). It is a complete and implementable (for example, with a compiler) description of the software. For more detail on the definition of open source software, see http://www.opensource.org/docs/definition.php.
18. The integrity of the FOSS development process is a prerequisite for effective shielding, as illustrated by the recent suit brought by the SCO Group against IBM. The suit alleges that IBM contributed SCO trade secrets to the Linux community. IBM has denied these allegations. Many observers, noting that Microsoft recently acquired an interest in SCO, see this as an attempt to undermine the reputation of the FOSS development process in general and the Linux development process in particular. Raymond and Landley (2003) offer a very useful if perhaps tendentious analysis of the suit.
19. Implications of this may soon be tested in court. GPL considerations have become prominent in the SCO Group suit against IBM. Allegations arose that SCO had incorporated Linux code into Unix System V in violation of the GPL. Other observers claimed that since SCO had itself shipped Linux containing the allegedly proprietary code, it had effectively GPL'd the code.
20. In some regions and countries (such as the European Union, Argentina, Brazil and Peru), governments have supported open source through procurement policies (for use in government ministries or departments). Germany has been especially Linux friendly. The federal government has even directly funded improvements to Linux user interfaces. In May 2003 the city of Munich decided to switch more than 14 000 desktop computers from Windows with Microsoft Office to Linux with Open Office. In June a migration of comparable size was announced in Britain. Other countries, such as Singapore, have provided tax breaks to companies that adopt open source products (Hahn 2002, ch. 1).
21. Game-theoretic analyses of FOSS projects include Bessen (2001), Harhoff et al. (2002) and Johnson (2002).
22. Harhoff et al. (2002) describe a number of interesting examples. For instance, Technicon Corporation devised automated blood chemistry analysers. The blueprint for this equipment built on the findings of laboratory clinicians who freely revealed them via publications. Technicon thus adopted the discoveries of lab clinicians without paying any royalty to them. The lab clinics were better off nonetheless because they induced the creation of more efficient equipment that analyses blood samples; Technicon is better off because the equipment is a commercial success. Another example is the Online Public Access Catalogues (OPACs), computerized systems that enable online access to a libraries collections. Vendor sharing of OPAC source code enabled users to suggest modifications to proprietary OPACs. Manufacturers eventually adopted them, again without paying royalties to the user innovators. The manufacturers had incentives to adopt them because enough users wanted the same thing. The users in turn got a better product.
23. The outrage of the FOSS community at such developments is important to consider when assessing explanations of their activities. If career signaling and peer recognition effects were the primary considerations of these contributors, why would such *ex post* developments be viscerally resented?

References

Allen, R.C. (1983), 'Collective invention', *Journal of Economic Behavior and Organization*, **4** (1), 1–24.

Bessen, J. (2001), 'Open source software: free provision of complex public goods', Research on Innovation Working Paper, <http://www.researchoninnovation.org/home.htm>, accessed 9 January 2004.

Bessen, J. and E. Maskin (2000), 'Sequential innovation, patents, and imitation', Working Paper 00-01, MIT Department of Economics, Cambridge, MA.

Boldrin, M. and D. Levine (2002), 'The case against intellectual property', *American Economic Review*, **92** (2), 209–12.

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

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

Cohen, Wesley, Richard Nelson and John Walsh (2000), 'Appropriability conditions and why firms patent and why they do not in the American manufacturing sector', NBER Working Paper 7552, National Bureau of Economic Research, Washington, DC.

Demsetz, H. (1967), 'Toward a theory of property rights', *American Economic Review*, **57** (2), 347–59.

Evans, D.S. and B. Reddy (2002), 'Government preferences for promoting open source software: a solution in search of a problem', Working Paper, National Economic Research Associates, Cambridge, MA.

Friedman, Lawrence M. (1985), *A History of American Law*, 2nd edn, New York: Simon & Schuster.

Hahn, Robert W. (2002), 'Government policy toward open source software: an overview', in Robert W. Hahn (ed.), *Government Policy toward Open Source Software*, Washington, DC: AEI-Brookings Joint Center for Regulatory Studies, pp. 1–11.

Hall, B.H. and R.H. Ziedonis (2001), 'The patent paradox revisited: an empirical study of patenting in the U.S. semiconductor industry, 1979–1995', *Rand Journal of Economics*, **32** (1), 101–28.

Harhoff, D., J. Henkel and E. von Hippel (2002), 'Profiting from voluntary information spillovers: how users benefit by freely revealing their innovations', Working Paper, MIT Sloan School of Management, Cambridge, MA.

Heller, M.A. (1998), 'The tragedy of the anticommons: property in the transition from Marx to markets', *Harvard Law Review*, **111** (3), January, 621–88.

Hirshleifer, J. (1971), 'The private and social value of information and the reward to inventive activity', *American Economic Review*, **61** (4), 561–74.

Johnson, J.P. (2002), 'Open source software: private provision of a public good', *Journal of Economics and Management Strategy*, **11** (4), 637–62.

Lemley, Mark, Peter Menell, Robert Merges and Pamela Samuelson (2003), *Software and Internet Law*, 2nd edn, New York: Aspen.

Lerner, J. and J. Tirole (2002), 'The simple economics of open source', *Journal of Industrial Economics*, **50** (2), 197–234.

Lerner, J., J. Tirole and M. Strojwas (2003), 'Cooperative marketing agreements between competitors: evidence from patent pools', NBER Working Paper 9680, National Bureau of Economic Research, Washington, DC.

Levin, R., A. Klevorick, R. Nelson and S. Winter (1987), 'Appropriating the returns from industrial research and development', *Brookings Papers on Economic Activity*, **3**, 784–829.

Machlup, Fritz (1958), *An Economic Review of the Patent System*, Washington, DC: United States Government Printing Office, <http://www.ipmall.fplc.edu/hosted_resources/jepson/unit1/aneconom.htm>, accessed 9 January 2004.

Machlup, F. and E. Penrose (1950), 'The patent controversy in the nineteenth century', *Journal of Economic History*, **10** (1), 1–29.

Mustonen, M. (2002), 'Why do firms support the development of substitute copyleft programs?', Department of Economics Working Paper, University of Helsinki.

O'Mahony, S. (2003), 'Guarding the commons: how community managed software projects

protect their work', *Research Policy*, Special Issue on Open Source Software Development, forthcoming.
Oracle Corporation (1994), 'Patent Policy', as reprinted by Gordon Irlam at <http://www.base.com/software-patents/statements/oracle.statement.html>, accessed 9 January 2004.
Organization for Economic Cooperation and Development (OECD) (1989), *Competition Policy and Intellectual Property Rights*, Paris: OECD.
Osterloh, M., R. Sandra and K. Bernhard (2003), 'Open source software production: climbing on the shoulders of giants', Working Paper, Institute for Research in Business Administration, University of Zurich.
Patent and Trademark Office (PTO), US Department of Commerce (1996), *Examination Guidelines for Computer-implemented Inventions*, Washington, DC, <http://www.uspto.gov/web/offices/com/sol/og/con/files/cons093.htm>, accessed 9 January 2004.
Raymond, E. and R. Landley (2003), 'OSI position paper on the *SCO vs. IBM* complaint', http://www.opensource.org/sco-vs-ibm.html, accessed 9 January 2004.
Romer, P. (1990), 'Endogenous technological change', *Journal of Political Economy*, **98** (part 2), S71–S102.
Rosenberg, Nathan (1976), *Perspectives on Technology*, Cambridge, UK: Cambridge University Press.
Schmidt, K. and M. Schnitzer (2003), 'Public subsidies for open source? Some economic policy issues of the software market', Discussion Paper 3793, Centre for Economic Policy Research, London, UK.
Shapiro, C. (2000), 'Navigating the patent thicket: cross licenses, patent pools, and standard setting', *Innovation Policy and the Economy*, **1**, 119–50.
Stallman, Richard (2000), 'The Anatomy of a Trivial Patent', <http://lpf.ai.mit.edu/Patents/anatomy-trivial-patent.txt>, accessed 9 January 2004.
Thomas, J.R. (2002), 'Liberty and property in the patent law', *Houston Law Review Symposium*, **39** (3), 569–619.
Torvalds, Linus (1998), 'Interview with Linus Torvalds: What motivates free software developers?', <http://www.firstmonday.dk/issues/issue3_3/torvalds/index.html>, accessed 9 January 2004.
von Hippel, Eric (1988), *Sources of Invention*, Oxford, UK: Oxford University Press.
Weber, Max (1923), *Wirtschaftsgeschichte*, New Brunswick, NJ: Transaction. References are to the 1981 reprint of 1927 translation, published by Greenberg as *General Economic History*.
World Intellectual Property Organization (WIPO) (1998), *Intellectual Property Reading Material*, Geneva: WIPO.

19 The governance of localized technological knowledge and the evolution of intellectual property rights

*Cristiano Antonelli**

Introduction

The economics of intellectual property rights has been characterized by an evolution that parallels and reflects the major shifts in the economics of technological knowledge that have occurred in recent years. This work provides an analysis of the effects of the changing foundations of the economics of knowledge upon the assessment of the design and the characteristics of intellectual property rights.

To do this, the chapter relies on a systemic approach to understanding the mechanisms of the institutional set-up that are most conducive to foster the rate of introduction of technological knowledge and hence technological change. The systemic analysis of the interdependent and complementary conditions of access and exclusion to the flows of technological interactions, transactions, coordination and communication that are specifically designed to organize the generation and the distribution of technological knowledge provides the appropriate context into which the role of each mechanism and specifically intellectual property rights can be assessed (Jaffe 2000, Nelson and Sampat 2001).

Major changes occurred in the economic understanding of knowledge in the second part of the twentieth century. Knowledge was first regarded as a typical public good that markets and profit-seeking agents could not produce in the appropriate quantities and with the appropriate characteristics. These theoretical ingredients paved the way to the build-up of the infrastructure for the public provision of knowledge. Consensus on the analysis of the public good character of knowledge was first contrasted and eventually replaced by the new argument about the quasi-private nature of technological knowledge. The identification of the central role of external knowledge in the production of new knowledge marks the third step. The identification of the knowledge trade-off stressed the limitations of intellectual property rights. Eventually a more balanced view based upon a deeper analysis of the generation and distribution of knowledge as a localized process and a closer analysis of the role of knowledge interactions and transactions as a part of a broader governance problem were elaborated. This evolution had important consequences

on the analysis and the implementation of institutional design for the organization of the production and distribution of knowledge (Machlup and Penrose 1950; Alchian and Demsetz 1973).

In the new analysis of the governance of the production and distribution of knowledge, intellectual property rights as signaling devices and the new understanding of the implications of knowledge as an essential facility play a major role.[1]

Knowledge as a public good

For a long time, the seminal contributions of Kenneth Arrow (1962) marked the economics of knowledge and provided the theoretical foundations for the build-up of public knowledge commons. In this approach, technological knowledge is seen as a public good because of its intrinsic limitations due to the high levels of indivisibility, non-excludability, non-appropriability and hence non-tradability. In this context markets are not able to provide the appropriate levels of knowledge because of the lack of incentives, and the missing opportunities for implementing the division of labor and hence achieving adequate levels of specialization.

The public provision of scientific and technological knowledge by means of funding to universities and other public research bodies, as well as directly to firms willing to undertake research programs of general interest, found a rationale in this argument. This led to the actual build-up and the systematic implementation of public knowledge commons.

The Arrovian approach impinged upon a second leg. The provision of public subsidies to firms undertaking research and development (R&D) activities was regarded as a necessary condition to remedy the low appropriability conditions and hence the lack of incentives.

Public procurement is the third basic tool to increase the production of knowledge. In particular, the demand for weapons becomes a major instrument to focus resources and identify research direction and objectives with a broader and general scope for derivative technological applications at the system level and relevant from the viewpoint of the general production of new scientific and technological knowledge. The natural leakage of technological knowledge from the military sector – often within the same corporations – feeds the levels of technological opportunity for the rest of the system. The spillover from the high-tech military activities provides unique opportunities for the introduction of product and process innovations in all other sectors of the economy.

The Arrovian approach easily integrated into the Schumpeterian legacy according to which the large corporation with substantial market power was the appropriate institution to accelerate the rate of introduction of technological change. Because of the low levels of natural appropriability, only large

incumbents in product markets characterized by barriers to entry, could fund R&D activities internally, with their own money. *Ex ante* monopolistic market power based upon barriers to entry in existing product markets would provide extra profits and hence secure the financial resources to fund R&D expenditures and, most importantly, reduce the risks of uncontrolled leakage and imitation. Competitors have yet to enter and entry is barred by substantial cost disadvantages. Appropriability is provided by barriers to entry rather than by barriers to imitation. The large corporation is also considered to be the appropriate tool to increase the rate of introduction of innovations as it provides internal markets for financial resources and competence: because of low appropriability regimes, arm's-length transactions in external markets cannot be used to coordinate either the allocation of financial resources into research activities and their selection, or the necessary division of scientific and technological labor.

The creation of intellectual property rights was regarded as a complementary institutional set-up, parallel to the public provision of scientific knowledge and the benign neglect of monopolistic market power. Patents and copyrights, if properly implemented, could reduce non-excludability and non-appropriability. In a proper institutional design, intellectual property rights may also favor tradability and hence lead to higher levels of specialization and division of labor in the technological applications of new scientific discoveries, made possible by public support. Intellectual property rights can help to increase the incentives to the production of incremental technological knowledge, but only in a broader context shaped by the role of the state (Kingston 2001).

Nevertheless, at this time intellectual property rights are not considered the major tool to improve the static and dynamic efficiency of the economic system in the production of knowledge. Patents are mainly viewed as an instrument designed to increase the incentives of firms to introduce minor technological innovations. Public subsidies, and public direct participation in the production and demand for knowledge are regarded as the basic instruments to push the introduction of radical technological innovations (Granstrand 1999).

Knowledge as a proprietary good

The first major shift in the economics of knowledge took place when the notion of knowledge as a public good was challenged and knowledge was regarded as a quasi-private good with higher levels of natural appropriability and exclusivity and hence tradability (Nelson and Winter 1982).

Technological knowledge is now viewed as the result of a bottom-up process of learning, which takes place mainly within the borders of firms. Technological knowledge is based upon tacit knowledge accumulated by means of a learning process. Eventually tacit knowledge can be articulated

and finally it translates into its codified form. Only when knowledge is fully codified and systematic articulation has been achieved, can it be diffused without the intentional assistance of the original holder (David 1993; Cowan and Foray 1997; Ancori et al. 2000; Cowan et al. 2000).

Imitation is hampered by major information and adaptation costs, appropriability is *de facto* secured by high levels of stickiness in routines and procedures: the 'not-invented-here' syndrome is much more effective than assumed in the public good tradition (Mansfield et al. 1981; Harabi 1995).

In the resource-based theory of the firm, the generation of technological knowledge is regarded as the distinctive feature of the firm. The firm does not equate with the production function and cannot be reduced to such a function because its essential role is the accumulation of competence, technological and organizational knowledge and the eventual introduction of technological and organizational innovations. From this viewpoint the firm precedes the production function: technology is the result of the accumulation of knowledge and its application to a specific economic activity (Penrose 1959; Foss 1997). The resource-based theory of the firm has grown as a development and an application of the economics of learning (Loasby 1999). It focuses on the characteristics of the process of accumulation of competence, the generation of technological knowledge and the introduction of technological and organizational innovations, not only as key factors to understanding the firm, but also as the relevant characteristics in the general production of technological knowledge.

In this context the firm is the primary actor in the production of knowledge for the whole economic system. The firm is viewed as the privileged locus where technological and organizational knowledge is generated by means of the integration of learning processes and formal R&D activities. The firm is considered primarily as a depository and a generator of competence and eventually knowledge (Foss 1997; Nooteboom 2000).

Because technological knowledge is now viewed as the sticky joint product of internal learning, it cannot spill freely into the air. Relevant absorption costs for potential users should be taken into account and qualified interactions between producers and users of new knowledge are necessary for technological knowledge to be actually transferred from one organization to another. The explicit and intentional assistance of original knowledge holders to prospective users is relevant, if not necessary.

The role of the public knowledge commons is questioned on two counts: first the firm is now viewed as the key actor in the production of knowledge and second, knowledge can circulate only if a dedicated framework of systematic interactions, which directly involve inventors, is put in place.

This new approach paved the way for significant steps towards the privatization of public knowledge commons (Argyres and Liebeskind 1998). The public provision of subsidies to firms undertaking R&D activities and the

direct role of the state in the production of knowledge are scrutinized more closely. The role of the university as the single provider of externalities to the economic system is questioned (Henderson et al. 1998).[2]

The new 'enclosures' replace the knowledge commons. Public research centers and universities were solicited to patent their discoveries and often forced to enter the markets for the technological outsourcing of large corporations. The conditions for the effective appropriation of knowledge are enforced both at the firm level and in public organizations: the mobility of human capital is increasingly regarded as a sensitive issue (May 2000; Cooper 2001; Mowery et al. 2001).

At the same time, the role of intellectual property rights is reconsidered. Intellectual property rights can complement and integrate the appropriability of technological knowledge, so that actual markets for knowledge, now much closer to traditional economic goods, can be developed. Intellectual property rights are now regarded as a complementary condition to increase tradability and consequently to achieve the standard conditions for an equilibrium supply of knowledge in the economic system. The extension of patent protection to new forms of knowledge such as software, algorithms and genetic entities finds its foundations here (Merges and Nelson 1994; Sakahibara and Bransletter 2001).

The discovery of the knowledge trade-off

The second major swing takes place when a closer analysis of knowledge appropriability made it possible to understand, next to its negative effects in terms of missing incentives and hence undersupply, the positive effects of technological spillover and the key role of technological externalities. The discovery of external knowledge, available not only through transactions in the markets for knowledge, but also through technological interactions, marks a new important step in the debate. External knowledge is an important input in the production process of new knowledge (David 1997). The appreciation of external knowledge as an essential input in the production of new knowledge, was later articulated in the systems of innovation approach, where the production of knowledge is viewed as the result of the cooperative behavior of agents undertaking complementary research activities (Eisenberg 1989; Scotchmer 1991; Autant-Bernard 2001).

The costs of exclusion associated with intellectual property rights, as a consequence, should be taken into account. Monopolistic control of relevant bits of knowledge, provided both *ex ante* and *ex post* by patents and barriers to entry in the products markets, respectively, can prevent not only its uncontrolled leakage and hence its dissemination but also further recombination, at least for a significant period (Arrow 1969; Dasgupta and David 1987, 1994; David 1993; Shavell and Ypersele 2001).[3]

The advantages of the intellectual property rights regime in terms of increased incentives to the market provision of technological knowledge, are now balanced by the costs in terms of delayed usage and incremental enrichment. The vertical and horizontal effects of indivisibility display their powerful effects in terms of cumulability. Indivisibility of knowledge translates into the basic cumulative complementarity among bits of knowledge. Complementarity and cumulability in turn imply that new bits of knowledge can be better introduced by building upon other bits already acquired, both in the same specific context and in other adjacent ones. The access exclusion from the knowledge already acquired reduces the prospect for new acquisitions and in any event has a strong social cost in terms of duplication expenses (O'Donoghue 2001).

The duration of exclusive property rights assigned by patents and the conditions for their renewal become a central issue for the possible negative drawbacks in slowing the rate of generation of new knowledge, especially with regard to general purpose knowledge with a wide scope of applications (Cornelli and Shankerman 2001; Scotchmer 2001; Shankerman and Scotchmer 2001).

The breadth of patents is also questioned: when it is large the protection is not specific and the negative effects in terms of foreclosure can easily exceed the advantages in terms of increased incentives. A narrow definition of the scope of application of intellectual property rights is thus recommended (Klemperer 1990; Merges and Nelson 1994; Hopenhayn and Mitchell 2001).

The introduction of a prize system has been advocated in this context as a possible alternative to patents. Prizes are seen as the proper incentive to the generation of technological knowledge because they combine the reward to innovators with informational advantages of patents in signaling the new relevant knowledge, which becomes available, but they do not impede the circulation of the new knowledge. The limitations of the prize system, however, become apparent in the screening and assignment procedure whereby a committee of scientists and technologists might easily assign the rewards to the wrong technological knowledge. An issue of bureaucratic coordination failure based upon bounded rationality clearly emerges (Wright 1983; Shavell and Ypersele 2001).

Here in the economics of technological knowledge the issues of externalities on the demand side become relevant and evident. The generation of technological knowledge is now characterized by technical and pecuniary externalities. The notion of user interdependence emerges when agents value the levels of usage of other agents of certain goods. As far as scientific and technological knowledge is concerned, interdependence among users, hence on the demand side, is very strong. The actual chances of generating a new relevant bit of knowledge for each agent depend upon the levels of accumula-

tion of skills and competence, education and access to information of the other agents in the community.

The issues of the distribution of knowledge become central in the debate and the notion of an actual knowledge trade-off is articulated. Uncontrolled leakage and low appropriability regimes reduce incentives, but may not necessarily lead to underprovision. Low appropriability engenders technological externalities and spillovers that are the prime factor in increasing the efficiency of generation of new knowledge, at the system level: the growth of efficiency can compensate for lower inputs (Griliches 1992).

Excess appropriability, both *ex ante* and *ex post*, based upon barriers to entry or on intellectual property rights, may slow down if not impede the working of knowledge complementarity, cumulability and fungibility. Intellectual property rights are now questioned as it seems evident that too strong a regime of protection may have positive effects in terms of increased incentives to the generation of knowledge, but has clearly negative effects in terms of delayed and slower circulation and distribution of the new knowledge available (Mazzoleni and Nelson 1998a and b).

The governance of the generation and distribution of localized technological knowledge

Further advance is made with the full appreciation of the notion of localized technological knowledge, which stresses the role of knowledge as a joint product of the economic and production activity. Agents learn how, when, where and what, also and mainly, out of their experience, accumulated in daily routines. The introduction of new technologies is heavily constrained by the amount of competence and experience accumulated by means of learning processes in specific technical and contextual procedures. Agents can generate new knowledge only in limited domains and fields where they have accumulated sufficient levels of competence and experience. A strong complementarity must be assumed between learning, as knowledge input, and other knowledge inputs such as R&D laboratories, within each firm (Antonelli 1999, 2004).

A second and most important complementarity takes place between internal and external knowledge. Firms can generate new knowledge and eventually introduce new technologies only when and if they are able to take advantage of external knowledge. No firm can rely exclusively on its own internal knowledge, either tacit or codified, whether it is the result of learning processes or formal R&D activities. The complementarity between external and internal knowledge and the cumulability between different vintages of knowledge, both internal and external, become key issues. Nor can firms generate new knowledge relying on only external or internal knowledge as input. Rather, with an appropriate ratio of internal to external knowledge, both

internal and external knowledge inputs enter into a multiplicative production function. Either above or below the threshold of the appropriate combination of the complementary inputs, the firm cannot achieve the maximum output.

Localized technological knowledge can be understood as a collective activity characterized by the complementarity both between external and internal knowledge and between the stock of existing knowledge and the flows of new knowledge. Markets appear to provide a unique set for incentive mechanisms to work swiftly: the result of such market interactions, however, may or may not lead the system towards stable and fair solutions. Tradability is a necessary but not sufficient condition for dynamic efficiency to be achieved. The aggregate outcomes of the governance mechanisms at the firm level are far from being attracted to a single equilibrium point.

Because of the complementarity between internal and external knowledge, especially if it is specified in terms of a multiplicative relationship, the aggregate outcome of both market transactions and interactions is unstable and sensitive to interactions and subjective decision making. When both demand and supply schedules are influenced by externalities, multiple equilibria exist (Marmolo 1999).

Inclusion needs to be coordinated and managed. Free riding can take place, although reciprocity and mutuality in interactions based upon knowledge barters, implemented by repeated and long-lasting exchanges, can help reduce the extent and the effect. Exclusion is dangerous because of the risk of missing the relevant complementary input, which characterizes the generation of new technologies.

The organization of the systems of innovation appears to be influenced by the need to implement and valorize the complementarity of the bits of knowledge possessed and accumulated in the diverse units, in a context characterized by relevant governance costs (Williamson 1985, 1996; Aghion and Tirole 1994; Garicano 2000).

The full identification of the notions of knowledge complexity, knowledge cumulability and knowledge fungibility is the major result of much empirical and theoretical work. The analysis of the intrinsic indivisibility of knowledge makes it possible to distinguish between (i) cumulability when it applies more precisely to the complementarity between the stocks of knowledge and the new flows, (ii) complexity when it applies to the variety of diverse elements of knowledge that are necessary to generate a new element of knowledge by means of recombination, and (iii) fungibility, when it consists of the variety of possible uses and applications of a given unit of knowledge that can be replicated with minimal incremental and variable costs.

The distinctive notions of knowledge transactions and interactions costs can also be identified and defined in terms of the costs of all the activities such as search, screening, processing and contracting that are necessary to

exchange bits of knowledge among independent parties. The trade-off between knowledge coordination costs, internal to firms, and knowledge transactions and interaction costs contributes to the understanding of the bundle of governance mechanisms at work (Antonelli 2001, 2003a, 2003b; Antonelli and Quéré 2002).

The analysis of the fabric of governance mechanisms of the production and distribution of scientific and technological knowledge emerges as the appropriate analytical framework. In the governance of knowledge not only is the traditional 'make or buy' trade-off relevant, but also a 'make or sell' choice has to be considered. The firm in fact needs to assess not only whether to rely upon external or internal knowledge in the production of new knowledge, but also whether to try and valorize the knowledge available internally as a good in itself and sell it disembodied in the markets for technological knowledge, or to use it as an input in the production of other goods. Technological strategies can be implemented by means of intentional learning procedures, internal R&D laboratories, technological outsourcing, location of R&D centers into technological districts, technological alliances and research joint ventures, and finally actual mergers and acquisition. Intellectual property rights play an important role within such a systemic context, together with other complementary and interdependent characteristics of economics systems such as the distribution of firms in regional space, the quality of financial markets and especially of the stock exchange, and needless to say, the organization of academic research (Dumont and Holmes 2002).

A wide range of choices in terms of governance can also be analysed and understood with respect to the characteristics of the processes of knowledge generation and usage. Different governance mechanisms and choices emerge according to the characteristics of technological knowledge and to the related levels of knowledge transaction costs (Dasgupta and David 1987a and b; Varsakelis 2001; Antonelli 2003a, 2003b).

The markets for knowledge: to sell or to make technological knowledge

Markets for technological knowledge are spreading in economic systems. The use of the marketplace to exchange technological knowledge is more and more common. Technological knowledge can be traded in a variety of ways: embodied in new intermediary and capital goods that enter into the production of other goods; as a knowledge-intensive business service; incorporated in weightless products such as software;[4] as a patent or a license; and finally, embodied in financial property rights after a new company has been created (Geroski 1995; Arora et al. 2001).

The characteristics of technological knowledge, its forms and the typology of the process by means of which new technological knowledge is generated, matter in assessing the appropriate mechanisms of governance and the weight

of knowledge transaction costs, that is, the costs for using the markets for technological knowledge.

Such knowledge transaction costs are relevant both on the demand and the supply sides. On the demand side, the identification of the agents holding specific bits of knowledge and the assessment of their quality is expensive in terms of search and screening costs, including the resources to evaluate the scope for incremental advance.

On the supply side, knowledge transaction costs arise mainly because of the high risks of opportunistic behavior of the customers. Uncontrolled usage of the knowledge can take place at the expense of the vendor. Derivative knowledge also matters: the vendor of the knowledge bears the risks of non-appropriation of the results of the implementation of the knowledge that has been sold (Scotchmer 1996).

The costs of writing proper contracts are relevant and there are a large variety of contingencies that must be taken into account. A strong intellectual property right regime and favorable conditions for its actual implementation in the markets for technological knowledge clearly favor the reduction of knowledge transaction costs. The role of the judiciary system with respect to the enforcement conditions of the contracts for disembodied technological knowledge is also relevant (Anand and Tarun 2000; Kingston 2001).

The main characteristics of knowledge identified so far are: (a) appropriability, defined in terms of the possibility of the inventor being the single beneficiary of the stream of profits associated with the introduction of a new bit of knowledge; (b) fungibility, defined by the scope of possible applications of a given unit of knowledge; (c) complexity, defined by the variety of complementary units of knowledge that are used to generate a new unit; (d) cumulability, defined by the vertical and diachronic complementarity between the stock of existing knowledge and the flow of new knowledge; (e) stickiness of knowledge in human capital and routines. These characteristics of knowledge have a direct bearing on its tradability, defined by the extent to which knowledge can be traded as a disembodied good in the marketplace.

The process by means of which technological knowledge is generated is also important in this context. Four different processes have been identified: learning, R&D, socialization and recombination. When recombination is the primary source of new knowledge and hence external codified knowledge matters, the access conditions to such knowledge are essential and intellectual property rights exert a key role. When socialization is important, that is the exchange of tacit knowledge in an informal context, qualified in terms of reciprocity and gift exchange, interaction in the knowledge communities is the primary vector. The social codes of conduct and the tacit laws of inclusion, exclusion and stratification in the knowledge community are the basic mechanisms of governance. When learning is the primary vector of tacit

knowledge and the latter is the primary source of new knowledge, the institutions of labor markets play a central role together with the organization of financial markets for the exchange of the property rights that embody the new knowledge in the form of new companies. When R&D activities, leading to new codified knowledge that cannot be easily appropriated, are the key means of generation of new knowledge, intellectual property rights are again in a central position.

The forms of technological knowledge are important: whether technological knowledge is more tacit, articulable or codified has a direct bearing on the governance of knowledge production. The exchange of tacit scientific and technological knowledge seems easier within research communities based upon repeated interactions and closed reciprocity in communication. Random inclusion can take place with positive effects, provided that newcomers are properly selected. The incentives to the creation of informal interaction procedures, often implemented by co-localization within technological districts, are very strong in this case. Geographical proximity emerges as a major factor conducive to closer knowledge interactions and exchanges: proximity reduces the scope for opportunistic behavior because of the exposure to repeated interactions and also reduces the costs of communication. Collective bodies such as industrial associations emerge as important governance structures, especially when technological knowledge is tacit and articulation requires complex procedures.

The exchange of articulable knowledge takes place more easily within vertical technological clubs and coalitions formed between vendors and customers–users. Vertical technological clubs differ from horizontal ones, where all parties are involved in a shared research activity. Vertical technological clubs complement the sale of patents and licenses and are based upon the close inspection of the activities of the customers and users of the patents. The relationship between the vendors and the customers takes place within long-term contracts, which include the assistance and the active cooperation of the two parties. The reputation of the members of the club plays an important role in building vertical technological clubs. The major goal here is the reduction of transaction costs stemming from the prospects for future knowledge: the vendors can retain the rights to participate in the appropriation of the derivative knowledge stemming from its implementation and incremental accumulation conducted by the customers. When technological knowledge is more articulable, the contractual interaction among partners within technological clubs can be better implemented. Here, knowledge transaction costs include high levels of monitoring and assessment of the actual conduct of the club members.[5]

Codified technological knowledge better meets the conditions for tradability when implemented by an appropriate intellectual property rights regime and

when the assistance of innovators is necessary and useful to reduce adoption and adaptation costs of perspective users. Codified knowledge is often found in fields where technological opportunities are slowing down and the levels of knowledge cumulability are lower (Cowan et al. 2000).

When technological knowledge can be easily appropriated by the innovator, either because of its complexity and hence natural levels of high appropriability, or because the regime of intellectual property rights is effective and easily enforced, firms may prefer to sell directly the technological knowledge as a good *per se* in the markets for knowledge.

On the contrary, with low levels of appropriability and hence low levels of tradability, firms cannot rely on the marketplace to valorize their intangible outputs. The embodiment of technological knowledge into new products and their eventual sale in the marketplace becomes necessary. The firm will choose to make and hence to include within the borders of the portfolio of activities the modules that use the knowledge as an intermediary input when the tradability and appropriability conditions are unfavorable. Here the governance choice for the firm is clearly between making and selling rather than between making and buying (Teece 1986, 2000; Antonelli 2001).

By the same token, the larger the cumulability of the technological knowledge, specific to the products and the production process of a firm, the larger the incentives for the internalization of the knowledge-generation process. The sale of technological knowledge in fact has high costs in terms of missed opportunities for further advances. The same argument applies when learning plays a key role in the generation of new knowledge: the full control of the production process is likely to yield important benefits in terms of increased rates of accumulation of new technological knowledge.

Knowledge fungibility has a direct bearing in this context. When the generation of new knowledge in operating downstream modules is directly influenced by the competence and the knowledge acquired in operating the module upstream, the firm has an incentive to make rather than to sell. This is true also when knowledge complexity applies and the operation of downstream modules has positive effects on the generation of new knowledge in the module upstream. Although the two modules are technically separated, high levels of indivisibility are found with respect to the generation of new technological knowledge and hence with respect to the introduction of new technologies.

When knowledge complexity and fungibility are weak, and knowledge transaction costs are low, the firm may choose to specialize: the modules are effectively separated both from the technical and the technological viewpoints.

Finally when both fungibility as well as transaction costs in the markets for technology are high, the firm has a strong incentive to use the technological

knowledge internally by means of downstream diversification in a wide range of products. When complexity and knowledge transaction costs are high, the firm has an incentive to integrate vertically in upstream activities.

Both downstream diversification and upstream integration in turn lead to an increase of coordination costs,[6] thereby inducing firms to sell knowledge in the marketplace, at least for a subset of applications. Here the costs of using both the market and the internal organization may be so high that the scope for a broad array of applications is lost. Information costs prevent firms from taking advantage of the full basket of technological opportunities stemming from the availability of technological knowledge with high levels of fungibility and complexity (Foray and Steinmueller 2003).

When technological knowledge is embedded in the learning routines of the firm, a strong intellectual property rights regime favors the use of financial markets as the appropriate governance mechanism. When technological knowledge cannot be separated from the organizational structures and human capital that characterize the localized learning process, the trade of the property rights of the company where the knowledge has been implemented becomes an effective mechanism that favors the division of intellectual labor as well as the distribution of knowledge and its appropriation (Gompers and Lerner 1999).

Patents as signals

The debates about the knowledge trade-off have concentrated on the positive and negative effects of the creation of intellectual property rights. Little attention has been paid to the informational role of such rights.

First, patents play a major role as signaling devices, which help the identification of the available bits of complementary knowledge and their owners so as to reduce search costs. Secrecy, the alternative to intellectual property rights, to secure exclusive ownership can have dramatic effects generally in terms of networking costs and specifically in the form of technological communication costs, and hence upon the number of knowledge complementarities that can be effectively activated (Oxley 1999; Teece 2000; Arundel 2001).

The appreciation of the informational role of patents has significant implications for their characteristics. With respect to the automatic granting of intellectual property rights, as in the case of copyrights, the selective and discretionary assignment of patents seems even more appropriate. The scrutiny of an authority is in fact most useful as a screening device, making it possible to sort out the bits of new knowledge that are actually relevant and useful. Similarly, patents assigned following the first-to-invent procedure is more useful than patents assigned with the first-to-file approach: the latter procedure characterizes the content of the patent in terms of novelty and ingenuity. It also seems clear that a narrow definition of the scope of a patent

is more useful, from an informational viewpoint, than a wide one. The identification and location of the relevant bits in the great map of knowledge become easier for each prospective user.

Second, intellectual property rights can provide not only a remedy for the public good nature of technological knowledge. They are also a remedy for tight vertical integration between the generation of new technological knowledge and its application to the production of new goods or to new production processes, rather than to its undersupply.

This analysis contrasts with the traditional argument, according to which the market supply of technological knowledge is deemed to undersupply because of its public good nature. The public good nature of technological knowledge does not necessarily lead to undersupply but rather pushes the knowledge-creating firm to use it as an intermediary input for the sequential production of economic goods. The markets for the products that are manufactured and delivered by means of the technological knowledge they embody can generate the incentives to produce the appropriate body of knowledge.

Effective property rights systems tend to favor the creation of markets for disembodied technological knowledge where the firms can specialize in the production of knowledge as a good *per se*. With a weak intellectual property rights regime, the holders of each bit of knowledge have a much stronger incentive to integrate vertically into the production of new goods and processes based upon innovative ideas, and to rely upon industrial secrets as a way of reducing informational leakage via a radical reduction in the circulation of the relevant bits of disembodied knowledge. The embodiment effect can be especially negative when the scope of application is wide and reverse engineering is complex, at least for unrelated prospective users.

Intellectual property rights reduce the incentive to internalize the valorization of technological knowledge by means of downward vertical integration. They can favor the creation of markets for technological knowledge and hence favor the distribution of fungible technological knowledge to a wider range of economic activities. But they do not necessarily increase the incentive to generate new knowledge, because of the sheer appropriability.

Third, the assignment of intellectual property rights seems by now a necessary condition not only to increase appropriability, but also as an institutional device which can improve the viability of the markets for knowledge and facilitate the interactions among holders of bits of complementary knowledge. Patents in fact can help transactions in the markets for knowledge because they make it easier for demand and supply to meet (Arora et al. 2001).

Following the resource-based theory of the firm, technological knowledge is the primary output of the firm: the firm exists because it is a depository of knowledge. The choice whether to sell it or to use it is especially relevant.

This approach can contribute to the debate on the economic organization for the supply of knowledge.

The new appreciation of the role of intellectual property rights is to be found in the assessment of their positive effects in terms of higher levels of specialization and division of labor. From this viewpoint the so-called 'knowledge trade-off', that is, the balanced assessment of both the positive effects of the monopolistic control of patents in terms of increased incentive to the supply of knowledge and the negative effects in terms of the reduced distribution of knowledge, needs to be reconsidered.

Because the systematic use of patents helps to identify bits of relevant knowledge for prospective users, it is essential to reducing the waste of duplication and facilitating the working of cumulability in the production of new knowledge. Patents can make knowledge interactions easier, provided that the exclusivity of ownership is properly tuned. The basic problems of the knowledge trade-off emerge again, but can be tackled in a different way (Pitkethly 2001).

Complementarities and property rights

The introduction and eventual implementation of new technological systems based upon new information and communication technologies have characterized the last decades of the twentieth century, with important implications for the governance of knowledge commons and the economics of property rights.

New information and communication technologies are characterized by the pervasive role of complementarities, with respect to the infrastructure and the goods produced and delivered. Moreover, high levels of complexity and interdependence also characterize the body of technological knowledge upon which new information and communication technologies build. Each advance in this field is strongly influenced by the access conditions for others, both diachronically and synchronically.

The telecommunications industry provides clear empirical and analytical evidence on such dynamics. Intrinsic complementarity among bits of networks as well as functions within the telecommunication networks is well known. Telecommunication networks are characterized on the supply side by the complementarity between existing trunks of each network and incremental portions: this leads to economies of incremental costs. Important complementarities, in terms of differentiated economies of density, also take place between functions within the network in terms of transmission, switching, distribution and signaling. Because of the differences in the minimum efficient size, the duration and the capacity of the different segments of the network, their combination into an integrated system yields important economies of scope. Finally, high levels of complementarity on the demand side,

with the well-known effects of network externalities characterize, telecommunication networks (Baumol et al. 1982).

The notion of essential facility has been elaborated on this basis. When a piece of property acquires the characteristics of an essential facility, the rights to access and interconnection cannot be exclusive. A separation between the rights of ownership and the rights of use is necessary in order for actual and workable competition to be implemented and eventually made possible (Baumol and Sidak 1994).

It is well known that competition in the telecommunications industry has been made possible by mandated interconnection. Mandated interconnection is a major factor of change and evolution in the definition of property rights. The ownership rights on the one hand and the rights of exclusive use on the other, traditionally associated in one single right, have been separated and rights of use of the network have been separated from the ownership rights. Firms do and can own telecommunication networks and can claim their property on all the segments of the network, but can no longer claim the right to exclusive usage. Other firms have the right to access the network and make a selective use of it. Dedicated authorities have been established since the late 1980s in most advanced countries in order to implement the right to interconnection, to regulate it and to fix the price of interconnection (Fransman 2002).

Communication authorities have been established to monitor the effective separation between the right of ownership and the rights of usage of telecommunication networks. Their activity here is essential because of the ever-changing conditions of the technology and hence the ever-changing conditions of the separation between ownership and usage. Second and most important, communications authorities have been established in order to fix *ex ante* the levels of interconnection tariffs. Interconnection tariffs must reflect properly the costs of the network and must make both appropriate returns on the investments for the owners as well as viable conditions of entry to new competitors. Newcomers must face actual competition in downstream markets with respect to incumbents and other competitors in the telecommunications industry (Madden 2003).

The evolution of property rights in the telecommunications industries has been the result of an understanding of the role of complementarities and the dual effects of economies of scope and externalities on the actual costs of both incumbents and new competitors in the industry. Mandated interconnection is indeed a significant departure from a fully-fledged and traditional definition of property rights.

A generalization can be made. The separation between ownership and rights of exclusive use is necessary within economic and physical systems where and when complementarities matter in order to restore and enforce the

conditions for competitive markets. The evolution of the property rights regime in the telecommunications industry is directly pertinent to the case of knowledge complementarities.

Indeed, knowledge shares all the relevant characteristics of an essential facility. Knowledge is characterized by intrinsic indivisibility and yet it is dispersed and fragmented in a variety of uses and possessed by a variety of owners. Each bit of knowledge is complementary to the next one along chains of weak and strong indivisibilities, which act both synchronically and diachronically. The exclusive access to each bit of knowledge can prevent others from cumulative undertakings.

The separation between ownership and usage conditions experienced in the case of the telecommunications industry can be applied successfully to intellectual property rights. The monopolistic rights given to inventors can, however, reduce the circulation of knowledge protected by intellectual property rights. Such effects are especially negative when knowledge complementarities apply and bits of knowledge can have important effects for the production of other knowledge in other fields of applications, often remote from those of the original invention and introduction.

The separation between the ownership of intellectual property and the right of exclusive use, already experienced with success in the telecommunications industry with the notion of mandated interconnection, can apply to this central and strategic area as well, where the reduction of the exclusivity of intellectual property rights can be realized by means of compulsory licensing and the liability rule. Compulsory licensing is increasingly associated with the authorization of mergers and acquisitions by antitrust authorities. Mergers are authorized provided that the firms agree to grant the licenses of their patents to all prospective users. The *ex ante* definition of the appropriate levels of the royalties can become a problem, however.

The transition towards the liability rule in intellectual property rights can be considered a useful device to implement mandated interconnection in intellectual property rights. Liability rule consists in the right of the owner of intellectual property to claim for appropriate payments for the usage of his/her rights (Kingston 2001).

In this context, the right of exclusive use is no longer associated with the rights of ownership of any intellectual property. As in telecommunication networks, ownership is recognized, as well as the right of other parties to take advantage of it for their own transmission needs.

In the case of intellectual property rights the *ex ante* definition of the equivalent of interconnection tariffs is questionable on many counts. First, research activities are characterized by high levels of risk and intrinsic uncertainty, both in terms of the chances of generating an output and with respect to the possible field of application of any innovation. *Ex ante* definition of the

costs of each new piece of knowledge is problematic. Much less difficult is the *ex post* identification of the economic value stemming from the application of a given specific piece of new knowledge.

The reduction of the rights of exclusive use of intellectual property, and the introduction of the mandated right to access intellectual property for third parties, combined with the eventual enforcement of the liability rule such that the judiciary system can help to secure *ex post* the payment of fair levels of royalties to the effective owners, can become an effective institutional innovation.

Intellectual property and hence patents can play a strong role in increasing the quality of the knowledge interactions. Full visibility of intellectual ownership can help to locate bits of complementary knowledge and hence reduce the costs of technological communication and networking activities in general, especially when the parties can reach agreement on the payment of appropriate royalties. By means of non-exclusive property rights, implemented by liability rules, knowledge interactions can come closer to market transactions and hence increase the scope for the valorization of knowledge complementarities.

Conclusions

Since the old days of knowledge as a public good, a lengthy process has been taking place. There is now a better understanding of the dynamics of knowledge accumulation, and appropriability conditions seem less relevant. Today, demand and network externalities play a much stronger role. Transactions in knowledge take place in markets characterized by knowledge transaction costs and governance mechanisms.

The better understanding of the generation of technological knowledge, made possible by the localized approach, and the results of the new enquiry in the economics of knowledge, draw attention to the economic characteristics of knowledge in terms of levels of fungibility, cumulability, complexity, stickiness and appropriability and its forms – tacit, articulated or codified. The analysis of the conditions for tradability is the ultimate result of all these advances. Tradability, however, is a not a sufficient condition for dynamic efficiency to be assured in the marketplace.

When increasing returns matter, as in the case of technological externalities, and the price mechanism is unable to convey all the relevant information, the markets are unable to set the right incentives and hence move in the right direction. Governance mechanisms at the microeconomic level and economic policy at the system level are necessary in order to provide the necessary coordination.

The systemic approach to understanding the mechanisms of the institutional set-up that are most conducive to fostering the rate of accumulation of

technological knowledge and its distribution, and hence the introduction of technological innovations, proves to be the appropriate analytical framework. The systemic analysis of the interdependent and complementary conditions of access and exclusion for the flow of technological interactions, transactions, coordination and communication that are specifically designed to organize the generation and the distribution of technological knowledge, emerges as an area of investigation and enquiry. All mechanisms and specifically intellectual property rights need to be assessed and considered within this broader framework.

The informational role of patents as carriers of relevant information about the actual levels of technological competence of agents and the availability of new bits of knowledge is increasingly appreciated (Stiglitz 2000).

The identification of each bit of complementary and useful knowledge as well as of the agents holding specific bits of knowledge and the assessment of their complementarity becomes an important function. This is expensive in terms of both search and opportunity costs: the costs of interacting with the wrong agents in terms of low opportunities. The selection of the firms and agents with whom technological cooperation and technological communication can take place is a relevant aspect of the governance mechanism and of the governance process. The creation of technological clubs and research joint ventures as institutional organizations designed to carry on collective research within selective coalitions can take place only if appropriate information is available on the technological competence of prospective members.

Technological signaling becomes relevant in this context as a device to reduce knowledge transaction costs. Patents are essential tools to signal the levels and the characteristics of the knowledge embodied in each organization. Patents are no longer regarded only as tools to increase appropriability, but also as devices to increase transparency in the knowledge markets and hence facilitate transactions.

This approach shows that intellectual property rights increase (i) the incentive to specialize in the generation of knowledge, (ii) the creation of markets for technological knowledge as a good in itself, and (iii) the production of knowledge.

A strong intellectual property rights regime associated with high levels of natural appropriability and codification and low levels of embeddness in the routines of the innovative firm can favor the use of markets as the appropriate governance mechanism to trade disembodied knowledge. When technological knowledge is sticky, embedded in learning process and organizational structures, thus making it difficult to trade as a disembodied good, the trade of the property rights of the company where the knowledge has been implemented becomes an effective mechanism which favors the division of intellectual labor as well as the distribution of knowledge and its appropriation.

A weak intellectual property rights regime favors the internal usage of technological knowledge within the borders of the corporation as an intermediary input. When *ex ante* and *ex post* appropriability is low, firms try to valorize technological knowledge as an intermediary input. When appropriability is high, firms may specialize in the direct generation and sale of technological knowledge. When technological knowledge has high levels of fungibility, and as such applies to a wide range of products and other technologies, a strong intellectual property rights regime may favor the distribution of technological knowledge. Vertical integration of technological knowledge with high levels of fungibility can lead to a reduced spectrum of applications because of rapidly increasing internal coordination costs.

The new assessment of the informational role of intellectual property rights in terms of increased incentives for the production and trade of knowledge, however, needs to be reconsidered, because of the perverse effects of exclusion on the efficiency of the generation of new knowledge, especially when radical innovations are at issue. The notion of knowledge as an essential facility becomes relevant. The extension and generalization of the notion of essential facility, elaborated in the telecommunications industry in the last decades of the twentieth century, is fruitful in the economics of knowledge and hence in the governance of knowledge commons. The evolution of the intellectual property rights regime towards the separation of ownership and the exclusive right of access to knowledge can provide important opportunities for the systematic valorization of both the markets for technology and the interactions among holders of complementary bits of knowledge. The mandated right of interconnection to bits of knowledge owned by third parties can take place with the implementation of the liability rule and the *ex post* payment of royalties without the preliminary consensus of the patent holders.

The reduction of exclusive rights in the use of intellectual property associated with the effective *ex post* enforcement of the liability rule can help the birth of the markets for knowledge. More efficient markets for knowledge can help to reduce the cost of interactions among complementary activities. Lower networking costs can increase the scope for the valorization of external knowledge complementarities. Easier access to external knowledge complementarities can increase the number of externalities and hence the levels of technological knowledge firms can rely upon. The final effect is clearly the generalized reduction of production costs and hence the increased levels of welfare.

Notes

* The funding of the research project 'Technological Knowledge and Localised Learning: What Perspectives for a European Policy?' under research contract no. HPSE-CT2001-00051 of the European Directorate for Research within the context of the Key Action

'Improving the socioeconomic knowledge base' is acknowledged, as well as comments and suggestions by Michel Quèrè and Martin Fransman.

1. A preliminary discussion along these lines can be found in Antonelli (2002) and in the comments of the author in Geuna et al. (2003).
2. A closer look at the working of the public commons and the actual need to scrutinize the productivity of the resources invested in the public knowledge commons, both at the system and the single units levels, is advocated (Jaffe and Lerner 2001).
3. The introduction of a prize system has been advocated in this context as a possible alternative to patents as the proper incentive for the generation of technological knowledge.
4. The case of numerical control provides the full range of cases. The technology of numerical control can be sold as a patent or a license. It can be sold embodied in software, in the numerical control itself or it can be embodied in a machine tool with numerical control. The machine tool in turn can be sold as such or it can be used as a capital good in the production of cars and trucks. The engineering industries, and specifically the packaging and textile machinery industries, provide similar evidence. The chemical industry is characterized by a similar trend with the identification of companies specializing in the supply of chemical plant design, as well as by companies that coordinate the competence in the design and delivery of the plants themselves. Finally, important companies in the chemical industries operate the full 'filière' of activities from the design of the plants, to their construction, to the production of market goods (Brusoni and Prencipe 2001).
5. The distinction between procedural and content contracts is relevant here. Procedural contracts are incomplete contracts designed to specify the modality of the interaction, while content contracts focus on the characteristics of the actual transaction. It is in fact possible to implement and eventually to enforce specific procedural contracts concerning the process of participation and timing of the assignment of property rights, temporary and partial exclusivity, time lags and partial and discriminated domains of privilege to subsets of contributors, selected according to both the amount of inputs and the actual results (Ménard 2000; Cassier and Foray 2002).
6. A notion of decreasing returns to scale in coordination activities with respect to the variety of modules seems plausible from an empirical viewpoint (see Chandler 1990; Argyres 1995).

References

Aghion, Philippe and Jean Tirole (1994), 'The management of innovation', *Quarterly Journal of Economics*, **109**, 1185–209.

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

Anand, N. Bharat and Khanna Tarun (2000), 'The structure of licensing contracts', *Journal of Industrial Economics*, **48**, 103–35.

Ancori, Bernard, Antoine Bureth and Patrick Cohendet (2000), 'The economics of knowledge: the debate about codification and tacit knowledge', *Industrial and Corporate Change*, **9**, 255–87.

Antonelli, Cristiano (1999*)*, *The Microdynamics of Technological Change*, London: Routledge.

Antonelli, Cristiano (2001), *The Microeconomics of Technological Systems*, Oxford: Oxford University Press.

Antonelli, Cristiano (2002), 'The economics of knowledge and the governance of knowledge commons', *Revista Brasileira de Inovacao*, **1**, 29–48.

Antonelli, Cristiano (2003a), *The Economics of Innovation, New Technologies and Structural Change*, London: Routledge.

Antonelli, Cristiano (2003b), 'Knowledge complementarity and fungibility: implications for regional strategy', *Regional Studies*, **39** (6–7), 595–606.

Antonelli, Cristiano (2004), "The economics of governance: the role of localized technological change in the interdependence among transaction, coordination and production', in Ken Green, Marcela Miozzo and Paul Dewick (eds), *Technology Knowledge and the Firm: Impli-*

cations for Strategy and Industrial Change, Cheltenham, UK and Northampton, MA, USA: Edward Elgar (forthcoming).
Antonelli, Cristiano and Michel Quéré (2002), 'The governance of interactive learning within innovation systems', *Urban Studies*, **39**, 1051–63.
Argyres, Nicholas S. (1995), 'Technology strategy governance structure and interdivisional coordination', *Journal of Economic Behavior and Organization*, **28**, 337–58.
Argyres, Nicholas S. and Julia P. Liebeskind (1998), 'Privatizing the intellectual commons: universities and the commercialization of biotechnology', *Journal of Economic Behavior and Organization*, **35**, 427–54.
Arora, Ashish, Andrea Fosfuri and Alfonso Gambardella (2001), *Markets for Technology*, Cambridge, MA: MIT Press.
Arrow, Kenneth J. (1962), 'Economic welfare and the allocation of resources for invention', in Richard R. Nelson (ed.), *The Rate and Direction of Inventive Activity: Economic and Social Factors*, Princeton, NJ: Princeton University Press for NBER, pp. 609–25.
Arrow, Kenneth, J. (1969), 'Classificatory notes on the production and transmission of technical knowledge', *American Economic Review*, **59**, 29–35.
Arundel, Anthony (2001), 'The relative effectiveness of patents and secrecy for appropriation', *Research Policy*, **30**, 611–24.
Autant-Bernard, Corinne (2001), 'Science and knowledge flows: evidence from the French case', *Research Policy*, **30**, 1069–78.
Baumol, William J., John C. Panzar and Robert D. Willig (1982), *Contestable Markets and the Theory of Industrial Structure*, New York: Harcourt Brace Jovanovich.
Baumol, William J. and John G. Sidak (1994), *Toward Competition in Local Telephony*, Cambridge, MA: MIT Press.
Brusoni, Stefano and Andrea Prencipe (2001), 'Unpacking the black box of modularity: technologies products and organizations', *Industrial and Corporate Change*, **10**, 179–205.
Cassier, Maurice and Dominique Foray (2002), 'Public knowledge, private property and the economics of high-tech consortia', *Economics of Innovation and New Technology*, **11**, 123–32.
Chandler, A.D. (1990), *Scale and Scope: The Dynamics of Industrial Capitalism*, Cambridge: The Belknap Press.
Cooper, David P. (2001), 'Innovation and reciprocal externalities: information transmission via job mobility', *Journal of Economic Behavior and Organization*, **45**, 403–25.
Cornelli, Francesca and Mark Shankerman (2001), 'Patent renewal and R&D incentives', *Rand Journal of Economics*, **32**, 107–213.
Cowan, Robin, Paul A. David and Dominique Foray (2000), 'The explicit economics of knowledge codification and tacitness', *Industrial and Corporate Change*, **9**, 211–53.
Cowan, Robin and Dominique Foray (1997), 'The economics of codification and the diffusion of knowledge', *Industrial and Corporate Change*, **6**, 595–622.
Dasgupta, Partha and Paul David (1987), 'Information disclosure and the economics of science and technology', in George Feiwel (ed.), *Arrow and the Ascent of Modern Economic Theory*, New York: New York University Press.
Dasgupta, Partha and Paul David (1994), 'Towards a new economics of science', *Research Policy*, **23**, 487–521.
David, P.A. (1993), 'Knowledge property and the system dynamics of technological change', in L. Summers and S. Shah (eds), *Proceedings of the World Bank Annual Conference on Development Economics*, Washington, DC: World Bank, pp. 215–48.
David, Paul A. (1997), 'From market magic to calypso science policy. A review of Terence Kealey's *The Economic Laws of Scientific Research*', *Research Policy*, **26**, 229–55.
Dumont, Beatrice and Peter Holmes (2002), 'The scope of intellectual property rights and their interface with competition law and policy: divergent paths to the same goal', *Economics of Innovation and New Technology*, **11**, 149–62.
Eisenberg, Rebecca S. (1989), 'Patents and the progress of science: exclusive rights and experimental use', *University of Chicago Law Review*, **56**, 1017–55.
Foray, Dominique and W. Edward Steinmueller (2003), 'On the economics of R&D and techno-

logical cooperation: insights and results from the project COLLINE', *Economics of Innovation and New Technology*, **12**, 77–92.
Foss, N. (1997), *Resources Firms and Strategies. A Reader in the Resource-based Perspective*, Oxford: Oxford University Press.
Fransman, Martin (2002), *Telecoms in the Internet Age: From Boom to Bust to …?*, Oxford: Oxford University Press.
Garicano, Louis (2000), 'Hierarchies and the organization of knowledge in production', *Journal of Political Economy*, **108**, 874–904.
Geroski, Paul (1995), 'Markets for technology', in Paul Stoneman (ed.), *Handbook of the Economics of Innovation and New Technology*, Oxford: Basil Blackwell, pp. 90–131.
Geuna, Aldo, Ammon Salter and W. Edward Steinmueller (eds) (2003), *Science and Innovation. Rethinking the Rationales for Funding and Governance*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Gompers, Paul and Josh Lerner (1999), *The Venture Capital Cycle*, Cambridge, MA: MIT Press.
Granstrand, Owe (1999), *The Economics and Management of Intellectual Property*, Amsterdam: Elsevier.
Griliches, Zvi (1992), 'The search for R&D spillovers', *Scandinavian Journal of Economics*, **94**, 29–47.
Harabi, Najib (1995), 'Appropriability of technical innovations: an empirical analysis', *Research Policy*, **24**, 981–92.
Henderson, Rebecca, Adam B. Jaffe and Manuel Trajtenberg (1998), 'Universities as a source of commercial technology: a detailed analysis of university patenting, 1965–88', *Review of Economics and Statistics*, **80**, 119–27.
Hopenhayn, Hugo A. and Matthew F. Mitchell (2001), 'Innovation variety and patent breadth', *Rand Journal of Economics*, **32**, 152–66.
Jaffe, Adam B. (2000), 'The U.S. patent system in transition: policy innovation and the innovation process', *Research Policy*, **29**, 531–57.
Jaffe, Adam B. and Josh Lerner (2001), 'Reinventing public R&D: patent policy and the commercialization of national laboratory technologies', *Rand Journal of Economics*, **32**, 167–98.
Kingston, William (2001), 'Innovation needs patents reforms', *Research Policy*, **30**, 403–23.
Klemperer, Paul (1990), 'How broad should the scope of patent protection be?', *Rand Journal of Economics*, **21**, 113–30.
Loasby, Brian J. (1999), *Knowledge, Institutions and Evolution in Economics*, London: Routledge.
Machlup, Fritz and Edith Penrose (1950), 'The patent controversy in the nineteenth century', *Journal of Economic History*, **10**, 1–20.
Madden, Gary (ed.) (2003), *Handbook on Telecommunications*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Mansfield, Edwin, Mark Schwartz and Samuel Wagner (1981), 'Imitation costs and patents: an empirical study', *Economic Journal*, **91**, 907–18.
Marmolo, Elisabetta (1999), 'A constitutional theory of public goods', *Journal of Economic Behavior and Organization*, **38**, 27–42.
May, Christopher (2000), *A Global Political Economy of Intellectual Property Rights. The New Enclosures?*, London: Routledge.
Mazzoleni, Roberto and Richard R. Nelson (1998a), 'Economic theories about the benefits and the costs of patents', *Journal of Economic Issues*, **32**, 1031–52.
Mazzoleni, Roberto and Richard R. Nelson (1998b), 'The benefits and the costs of strong patent protection: a contribution to the current debate', *Research Policy*, **27**, 273–84.
Ménard, Claude (ed.) (2000), *Institutions Contracts and Organizations: Perspectives from New Institutional Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Merges, Robert and Richard R. Nelson (1994), 'On limiting or encouraging rivalry in technical progress: the effect of patent scope decisions', *Journal of Economic Behavior and Organization*, **25**, 1–24.
Mowery, David C., Richard R. Nelson, Bhaven N. Sampat and Arvidis A. Ziedonis (2001), 'The

growth of patenting and licensing by U.S. universities: an assessment of the effects of the Bayh–Dole Act of 1980', *Research Policy*, **30**, 99–119.
Nelson, Richard R. and Bhaven N. Sampat (2001), 'Making sense of institutions as a factor shaping economic performance', *Journal of Economic Behavior and Organization*, **44**, 31–54.
Nelson, Richard R. and Sydney G. Winter (1982), *An Evolutionary Theory of Economic Change*, Cambridge, MA: Harvard University Press,
Nooteboom, Bart (2000), *Learning and Innovation in Organizations and Economics*, Oxford: Oxford University Press.
O'Donoghue, Ted (2001), 'A patentability requirement for sequential innovation', *Rand Journal of Economics*, **32**, 654–79.
Oxley, Joanne E. (1999), 'Institutional environment and the mechanisms of governance: the impact of intellectual property protection on the structure of inter-firm alliances', *Journal of Economic Behavior and Organization*, **38**, 283–309.
Penrose, Edith T. (1959), *The Theory of the Growth of the Firm*, Oxford: Basil Blackwell.
Pitkethly, Robert H. (2001), 'Intellectual property strategy in Japanese and UK companies: patent licensing decisions and learning opportunities', *Research Policy*, **30**, 425–42.
Sakahibara, Mariko and Lee Bransletter (2001), 'Do stronger patents induce more innovation? Evidence from the 1988 Japanese patent law reform', *Rand Journal of Economics*, **32**, 77–100.
Scotchmer, Suzanne (1991), 'Standing on the shoulders of giants: cumulative research and the patent law', *Journal of Economic Perspectives*, **5**, 29–41.
Scotchmer, Suzanne (1996), 'Protecting early innovators: should second-generation products be patentable', *Rand Journal of Economics*, **27**, 322–31.
Scotchmer, Suzanne (2001), 'On the optimality of the patent renewal system', *Rand Journal of Economics*, **32**, 181–96.
Shankerman, Mark and Suzanne Scotchmer (2001), 'Damages and injunctions in protecting intellectual property', *Rand Journal of Economics*, **32**, 199–220.
Shavell, Steven and Tanguy Van Ypersele (2001), Rewards versus intellectual property rights, *Journal of Law and Economics*, **44**, 525–47.
Stiglitz, Joseph (2000), 'The contributions of the economics of information twentieth century economics', *Quarterly Journal of Economics*, **115**, 1441–78.
Teece, D.J. (1986), 'Profiting from technological innovation: implications for integration collaboration licensing and public policy', *Research Policy*, **15**, 285–305.
Teece, David J. (2000), *Managing Intellectual Capital*, Oxford: Oxford University Press.
Varsakelis, Nikos C. (2001), 'The impact of patent protection, economy openness and national culture on R&D investment: across-country empirical investigation', *Research Policy*, **30**, 1050–68.
Williamson, Oliver E. (1985), *The Economic Institutions of Capitalism*, New York: Free Press.
Williamson, Oliver E. (1996), *The Mechanisms of Governance*, Oxford: Oxford University Press.
Wright, Brian D. (1983), 'The economics of invention incentives: patents, prizes and research contracts', *American Economic Review*, **73**, 691–707.

20 Property rights in the digital space

*Eric Brousseau**

Internet, a global and integrated information space

Digital technologies overwhelm the economics of information, knowledge and networks. First, they increase the fixed-costs nature of these resources, turning them into less rival goods than before (Shapiro and Varian 1999). Second, the digital codification of information allows the separation of the management of containers from the management of contents, leading to 'universal' platforms able to manage any kind of information independent of its nature (whether it is information, codified or tacit knowledge), its form (voice, image, data and so on), or its semantics (whatever 'language' is used to establish links between things, concepts and form of perceptible expression). When compatible technical solutions are implemented across groups of users, the platform becomes global. It enables any agent to transmit any information to any third party or to access contents. Third, the rules that govern the use of information can be implemented in the software that manages the hardware. This provides the opportunity to implement self-enforcing rules about the possible use of the technology and the information (Lessig 1999).

Of course, none of these characteristics is perfect in the digital world. Variable costs are not equal to zero. Technical standards are competing and imperfect. Hackers constantly break codes. Moreover, these characteristics do not free economic agents of all constraints. Solutions have to be implemented to cover fixed costs and to stimulate agents to contribute to the production of public goods. Coordination mechanisms have to (be) develop(ed) to ensure the development (and enforcement) of compatible and efficient technical standards. Rules have to be designed and selected to enhance the efficiency of the information and knowledge-based economy. At the same time, digital technologies give rise to many economic and institutional questions, and their specific characteristics lead one to wonder what the optimal way to organize the institutional frameworks of the digital space would be.

The property right (PR) system is one of the essential components of an institutional frame. As pointed out by Barzel (1989) and North (1990), by establishing how agents can use and exchange rights to make decisions about the use of resources, a PR system affects the way an economic system performs. In this chapter, we shall point out how digital technologies call for the emergence of *ad hoc* frameworks to organize the production and the use of information and knowledge on digital networks.

We shall focus on the Internet since it is not only the current most famous communication and information-sharing network, but also, and essentially, the integration platform of digital technologies. Indeed, the Internet is not a network but a set of standards (and an addressing system) that enables heterogeneous information processing devices (IPDs) – either computers or any system able to code/decode, store/retrieve, receive/transmit or process information – to exchange information and to process it cooperatively. Many of the limits of the present Internet – low speed (it is unable to transmit high definition moving images efficiently), unreliability (some exchanges of information cannot be completed when the network is crowded), security risks (information can be captured when transiting through the network, disk contents can be accessed by third parties (hackers) and so on) – should be removed in the future, thanks to the development of the technology and the institutional framework that will set up rules to manage networks and information. The Internet will then become the principal platform for the production, circulation and consumption of knowledge and information.

Stakes are indeed huge. Technology provides new opportunities to organize an institutional environment able to fully benefit from the capability to manage information and knowledge more efficiently. In addition, the Internet is challenging the current institutional frameworks. There are, however, multiple options to build the new one. There are at least two highly controversial issues. First, should the new institutions come from the existing national states – on the model of intergovernmental organizations – or should they be based on self-regulation? Second, should the new institutions maintain the current legal segmentation following the nature and the form of information (for example, copyright versus patents, contrasts with regard to legislation of the press and of entertainment activities, differences in the protection of music and movies and so on), or should they be homogeneous? Indeed, once digitized, any type of information can be managed the same ways on digital networks. These issues are too broad to be discussed in just one chapter. Here we shall therefore limit our focus to the economics of intellectual property rights (IPRs) on the Internet.

We shall argue that, while digital technologies make it possible to establish a decentralized IPR system based on self-regulation and the self-implementation of exclusive rights of use over information, a total decentralization would not be optimal. On the one hand, decentralization would enable agents to benefit from coordination frameworks well adapted to their specific needs and preferences. On the other hand, full decentralization of the settlement of IPRs would result in inefficiencies. While becoming subject to exclusion, most information remains a non-divisible good. Individuals and groups could succeed in establishing monopolies that would deter further entries into 'privatized' information spaces. In addition, despite the capabilities of the

technology, it remains individually and collectively costly to enforce exclusive rights of use. Full decentralization could therefore lead to prohibitive transaction costs. Several elements are therefore calling for a coordination of self-regulatory efforts to settle IPRs. This coordination should be organized by a central entity in charge of promoting the collective interest by reinforcing the ability of individuals and community to self-organize, by preventing regulatory overlaps and inconsistencies, by maintaining the sustainability of competition in the long run and by taking into account the specificity of non-rival goods. This central entity should be of a federal nature, and should coordinate the efforts of the various self-regulators (whether groups or individuals) to ensure the consistency and the efficiency of the global digital space.

First, we shall explain how the PR approach developed by Barzel (1989) and North (1990) makes it possible to analyse the organization of the institutional framework within which agents can use and exchange economic resources. We shall then illustrate how and why digital technologies and networks challenge the 'traditional' institutional frameworks that organize access and use of information goods. The problems raised by the management of rival and non-rival information goods will then be examined. These will lead us to explain the case for a federal system to coordinate self-regulatory entities, so as to ensure the consistency and efficiency of the exchange of digital goods.

The transactional approach to property rights

According to Barzel (1989) and North (1990) a PR system is a set of rules and mechanisms that delineates rights over economic resources and allocates them to decision makers so as to enable them to take economic actions.[1] It is based on a definition of these rights, consisting in setting the frontiers among different ways of using resources and among regimes for appropriating the output of these uses, and on a process of allocation of these rights, which are granted to individuals or groups. These operations are qualified as 'measurement' by Barzel and they generate measurement costs. Enforcement mechanisms implement these rights of use by excluding every unentitled agent from access to the protected resources, or from capturing the benefits. This implies controlling access, supervising uses, granting authorizations for uses and punishing unauthorized uses (either to obtain compensation for damage or to deter potential infringers), and generates enforcement costs.

In a given group – say, a nation – measuring and enforcement of property rights can be performed either centrally by an authority of last resort – generally the state, which exercises the monopoly of legitimate violence – or by the agents. In the former case, the government defines for each set of economic resources the rights associated to them (for example, *usus*, *fructus*

and *abusus*) and maintains a registry where each of these rights is attributed to individuals or groups. Then the government sets up and operates an enforcement mechanism to expel any unentitled agents from the protected use of these resources. It can be an *ex ante* mechanism, for example, a guard or an encryption mechanism that forbids access, or an *ex post* mechanism that assesses violation and punishes infringers. The alternative is to have the property rights self-delineated and self-enforced by agents. In this latter case, individuals (or groups) claim exclusive usage, and they apply all the available means (and in the last resort violence) to have their claims enforced by third parties.

The advantages of centralization are threefold (Barzel 1989; North 1990; Bessy and Brousseau 1998; Brousseau and Fares 2000). First, it allows economies of scale and scope, as well as learning effects in the measurement and enforcement of operations. Indeed, the centralized establishment of rights avoids duplication of efforts and enables individuals to specialize. Second, it reduces the level of conflict since agents are less likely to adopt predatory strategies.[2] On the one hand, if a central authority defines and allocates rights of uses, individual agents cannot unilaterally expand the boundaries of their rights to the detriment of others. On the other hand, if a central authority enforces the existing rights, incentives to infringe them are reduced since it is either *ex ante* or *ex post* costly to do so. Third, the central authority can limit the room for manoeuvre of agents and thus reach a better collective outcome. Decentralized implementation processes could indeed lead to collective inefficiencies when externalities occur.[3] In the case of non-rival goods, the central authority could bind the agents' ability to capture them, so as to maximize their spread or their use.[4] More generally, when a monopoly position generates inefficient capture,[5] it can be optimal to have an authority that prevents such inefficiencies. Lastly, a last resort intervention can be efficient in the case of negative externalities. Think for instance of the enforcement of exclusive rights of use. The existence of a last resort authority prevents agents both from overequipping themselves in attack and defence capabilities (since decentralized enforcement would lead to an 'arms race' to be always able to exercise credible threats among others, while being non-subject to extortion by stronger players), and from expending resources in destructive conflicts. Of course, what has been said depends upon the assumption that the central authority is both efficient and benevolent.

The cost of centralization is twofold. First, it leads to inefficiencies due to inadequate adaptation of central uniform solutions to local diversified needs and preferences. Even if the centrally designed rule can be optional, it is obvious that it cannot take into account all the possible options. This would require unbounded computing capabilities and unlimited access to information by the central authority (for example, it would be able to access the set of

preferences of each individual). Moreover, if it were feasible, the benefit of centralization would be lost since economies of scale and learning effects would be dissipated in the design of rules adapted to any specific individual and to any particular situation. Second, centralization leads to inefficiencies, since it induces distortion between the marginal cost of delimiting rights over the various uses of a resource and the marginal benefits of doing it. Indeed, delimiting and enforcing property rights have a cost. Since the central authority benefits from economies of scale, scope and learning, it cannot charge any individual benefiting from a right to use a resource according to the marginal benefit he/she gets from it. Moreover it would be tremendously costly to try to evaluate the marginal cost of any protection and the marginal individual benefit. Public and centralized PR systems are therefore financed by taxes (or fixed fees). These taxes (and fees) are distortive by themselves. They are also distortive since the government could dedicate a lot of resources to delineate exclusive rights of uses of little economic value; while at the same time it would underprotect access to other more valuable uses of resources.

The advantages of decentralization are the opposite. Decentralization allows a finely adapted definition of property rights to the preferences of economic agents, and ensures that only the use of resources that generate a utility whose valuation is over the cost of protection will be protected. This guarantees, first, that effort will not be dedicated to design and enforce property rights of poor economic value, and second, that certain uses of some resources will remain freely available, which can raise efficiency if these resources are indivisible and renewable. The costs are also the opposite. Decentralization is costly for the agents because they have to bear the direct costs of measuring and enforcing the exclusive rights they claim (without benefiting from economies of scale, scope and learning). Moreover, decentralization is the cause of many and permanent conflicts since there is no last resort authority to stop them. There can be both conflicting claims and conflicting enforcement of these (never recognized) exclusive rights of use. Agents will therefore overinvest in attack and defence capabilities and the high level of risk will prevent part of the uses and the trade from happening (because insurance premiums would be too high).

Decentralization also has a dynamic advantage. It facilitates individual innovation with regard to norms design, resulting in an institutional framework that is more dynamic and more able to adapt than when innovation is centralized. In particular, local norms that are particularly efficient would be locally settled by individuals and progressively adopted by the members of an emerging community even if there is no social consensus about them. In a sense this is what happened with the norms that govern open-source communities. The boundaries of that decentralized innovation process are obvious. First they can result in inconsistencies. Second, since competition can be

biased, efficient norms do not systematically prevail or survive (as pointed out by the literature on technology diffusion; for example, David 1985).

In practice, setting up a PR system, either centrally or decentrally, would be inefficient in the sense that the costs of setting up a complete PR system[6] would be too high as compared to the benefit agents would get from being able to use resources, investing in production capabilities, and organizing trade.

Any PR system results therefore from a trade-off between the advantages of centralization and those of decentralization. The central authority designs an incomplete PR system, and the agents decentrally complete it (or vice versa). From a normative[7] point of view, this trade-off should make it possible to maximize the 'collective welfare/transaction costs' ratio; the latter notions encompassing the cost of resources dedicated to trading and setting up the PR system (either centrally or decentrally). It would then be dependent on three main factors: the nature of the resources (which are subject or not to many different uses by many different agents), the heterogeneity of preferences of agents, and the capabilities of the central institutions. This is why the organization of PR systems differs across space and time, given the nature of the resources. This trade-off results in a level of transaction costs (at the macro level), and a distribution of them (at the micro level), and impacts on the capability of a system to generate wealth.

The Internet as a challenge to traditional institutional frameworks

A global information infrastructure overwhelming the economics of information goods

Digital technologies raise essential issues with regard to IPRs since these technologies affect goods that (often) have a public nature, and whose circulation can be organized on a global basis, while new institutional frameworks can be imagined since the technology allows the design of self-enforcing rules. More precisely:

- The Internet will become the principal infrastructure for exchanging and sharing information and knowledge. It can be information and knowledge that is exchanged *per se*, or which are themselves components of modular goods and services that mix tangible and intangible components, or even information that is necessary to organize a transaction. Part of this information is clearly a public good. Part of it is private, since it is submitted to rival uses. Given the nature of information, the optimal way of organizing rights of access and rights of use differs, resulting in complex institutional frameworks.
- The Internet is organizing a global connectivity that is both one of its ends and one of the means of benefiting from a reliable and evolving

network infrastructure (since new functionalities can be incorporated into the network only by implementing new IPDs or new software). This connectivity is provided by universal technical standards that manage interoperability among network components.[8] Interface standards enable any device connected to the Internet to exchange and cooperatively process information with another device implementing the same standards. The strength of the Internet standards relies in their width, openness and public nature. They are wide in the sense that they organize interfaces among a very wide set of technologies and software. They are open in the sense that they are modular and are permanently enhanced to ensure interoperability among most of the available information technologies worldwide. They are public in the sense that they have been produced on the open-source software model and are available for free to any user or developer. Due to these characteristics, they have been able to generate high positive externalities of adoption. They ended in creating a global network that is overwhelming most pre-existing information gaps between individuals and professionals, between large and small firms, among economics agents involved in different industries, among citizens in different countries and so on.

- The Internet is based on a decentralized architecture, qualified as 'end to end'.[9] The IPDs connected to the network can get in touch directly with the other machines without depending upon any central capability that would manage the network. This decentralization of the network administration allows any user of the Internet to freely organize information space by establishing technical rules designing how a set of machines can interoperate to share or to exchange information or processing capabilities. This can be done freely because end to end imposes no constraint on how to do it and because there are no means of preventing users from organizing such spaces. This provides the Internet users with the ability to decentrally set up the rules that will govern their information spaces.

Digital technologies provide new opportunities to settle rules about the uses of information resources, while the traditional institutional frameworks organizing the management of these resources – the national IPR systems, but also most national regulations on contents – are both outflanked and severely questioned by the rise of a transnational and universal infrastructure. In particular, digital technologies make it possible to implement self-regulations at a low cost, which can provide the members of communities (characterized by common preferences, or common interests) with rules that better fit their needs than when general rules are designed to govern heterogeneous communities.

Code and controlled information spaces as means of self-regulation

The combination of the code and end-to-end connectivity makes it possible to implement a decentralized process of self-enforcing regulations in the cyberworld.

As pointed out by Lessig (1999), writing digital codes is equivalent to writing rules. Two techniques are at the heart of this feature. First, software codes implement routines on how a set of information will be handled. Second, encryption capabilities allow the control of access both to information and to software. By combining encryption and software coding an agent is able to control access and use of digitized information goods and services. He/she is therefore able to delineate rights of access and rights of uses on information goods and services – and also on network components – and to grant them to any third party.

These rights are self-enforceable since the simple fact that it is written in code makes it mandatory to use the information the way it is authorized or imposed by the code. Of course there are limits to this self-enforcement since a code can be cracked. However, cracking a code requires expertise and time. It is a costly activity. Most of the individuals or groups able to crack codes perform, even informally, a cost–benefit analysis to determine whether or not it is relevant to bear the fixed costs of code cracking. The result of the analysis will depend upon the benefits that are a function, on the one hand, of the conditions of access to the coded contents (both depending on the tariffs and the restrictions imposed by the writer of the code), and on the other hand, of the rewards the cracker can get either by selling or by disclosing for free the 'uncoded' material (which itself is dependent upon what the hacker can get – payment, reputation and so on – and the size of the community). The possibility of cracking codes makes the self-enforcement of rules implemented in digital codes imperfect. The ability to control the uses of information thanks to digital technologies is nevertheless quite strong for at least three reasons. First, a code has to be written to manage hardware in any case. Cracking *per se* is useless if no code is written to replace the cracked code. Many codes are not cracked simply because nobody wants to pay for writing a new one. Second, due to their network nature, information technologies perform in a system and compatible codes have to be adopted. Third, the new technological base enables the control of uses not only *ex ante*, but also *ex post*. Thanks to the low cost of handling and storing information, and since any operation requires the execution of codes, a systematic tracking of information handling operations is performed by most digital information systems. Infringements to rules can easily be tracked and then retaliations can possibly be implemented. This ability to design (almost) self-enforceable rules about the way information can be used, strongly reduce the usefulness of having such rules designed and enforced by a third party of last resort, such as the state.[10]

The end-to-end connectivity also plays a central role, in enabling the agents to design and enforce rules without the intervention of a third party. Indeed, it enables agents to organize the information space, whose frontiers can be controlled. Thanks to end to end, agents can control who (or which machine) can access or not a virtual space within which the participants can communicate, share information, perform cooperative information handling processes and so on. In concrete terms these virtual spaces can be websites with controlled access, intranets or extranets, mailing lists and so on. This ability to control inclusion in/exclusion from virtual space allows, first, the setting of frontiers within which common rules are to apply, second, having these rules enforced, since the ability to exclude provides the agent(s) in charge of managing the virtual space with means of retaliation. The credibility of these retaliations is obviously bounded by the (implicit) cost–benefit analysis made by agents accessing an information space. On the one hand, access provides various possible advantages: lower transaction costs among members of the community (see Milgrom et al. 1990), free access to shared information and more generally to a club-good and so on. On the other hand, it can be costly to be excluded, especially if sunk investments were requested to join. The higher the advantages, the higher the sunk costs, the fewer alternatives to the considered information spaces, the higher the credibility of potential retaliations. While enforcement capabilities are partly bounded, the end-to-end connectivity provides the agents with the ability to implement self-regulations in the digital world. Indeed, they can create information spaces with clear boundaries and decide that the infringers of common rules, whether they are unilaterally or consensually identified, will be expelled from the space, and therefore from the virtual community it creates.

This is typically what happens in many 'virtual' communities on the Internet, among which those of open-source software developers are the most famous. When joining a 'project' – like Linux, Apache, Mozilla and so on – developers gain access to a source code (which is hidden in commercial software). The source code enables the user to understand how the software operates. It then allows the user either to enhance it, or to add new functionalities to the software. The GNU licences that are at the core of these virtual communities stipulate that in exchange for free access to the code, developers have to disclose their own lines of source code. Several retaliation means can be implemented if there are infringements to these rules. In particular, the opportunistic developer can be prevented from further access to the source code, can be subject to ostracism and even to retaliations (spam, viruses). The same applies in many forums, discussion lists, chat rooms and so on. These apparent anecdotal practices of techno-alcoholics are also used by many business organizations to structure information sharing within their Intranet. It is also a basis of the organization of markets.

Thus, digital technologies provide individuals or groups with an accessible tool to design rules and to have them enforced. These rules can be individually settled and concern the way information and knowledge[11] can be used. Individual agents are able to implement at a relatively low cost – the cost of writing a digital sequence to manage authorization of access, plus (possibly) the costs of tracking uses – rights of access and uses. Moreover, these rights can be traded since they are implemented in the set of digits that is potentially transmitted among information processing devices. When a rights holder transmits a digital sequence to a third party, he/she can implement in the sequence the contractual conditions in which the receiver can access and use the content. This contract is self-enforceable since the code controls *ex ante* the future uses. A system of tradable rights of uses can therefore be implemented without any recourse to a central institution. In addition, individuals and groups have the possibility of creating information spaces in which they settle rules that have to be enforced by the members of the community. These rules cover the use and access to information and knowledge, but they could be even more general since, for instance, an information space can be the 'information infrastructure' of a market. In that case, the rules that will organize a community do not create rights of uses only on information goods.

The discrepancy between the global and generic information infrastructure and traditional regulatory frameworks

Traditional institutional frameworks are therefore challenged by digital technologies. Individuals and groups can indeed create and implement *ex nihilo* property rights, contracts and exchange rules, and regulations bounding the extent of these property rights. Moreover, digital technologies weaken these traditional institutional frameworks, since they enable agents to bypass them.

The Internet is a-territorial by nature, while public legal systems are implemented on a territorial basis. The Internet's generalized interconnection and decentralized management provides individuals with the ability to easily manipulate information at a low cost and to use it according to specific rules in information spaces that do not fit the territories of jurisdiction, and/or that can escape the sovereignty of enforcement authorities. These information spaces defined by on-line communities are generally international. Conflicting legal principles should therefore often apply. Moreover, digital networks can support uses that are hybrid as compared to the pre-existing categorization of uses. Think, for instances of chats, forums and discussion lists that correspond neither to program broadcasting, nor to pure interpersonal communication. This is another reason for the existence of potential conflicting legislation. The discrepancy among legal spaces and information spaces makes it difficult to apply legal rules.

First, it is often difficult to determine which law should apply. In many cases conflicting laws could legitimately apply, and there is no pre-existing international convention to solve potential conflicts among laws. Moreover, digital technologies make it possible to distribute the processing of information through the network. A well-designed information service might locate the various components of an information handling process in computers under different jurisdiction, which could perform illegal information handling processes according to a national law, without formally breaking laws. Lastly, in many cases, existing legal rules have not been translated to be applicable to the new media, leaving open wide spaces free of law.

Second, enforcement authorities are often unable to act since authors of legal infringements are difficult to identify and beyond the reach of the authorities' power of sanction:[12]

- Infringements and infringers are not so easy to identify. It would be complex for a governmental agency to efficiently supervise the exchanges of information among citizens (or the organization that acts under their jurisdiction) and between them and foreign third parties to guarantee the enforcement of existing laws. Moreover, such a systematic supervision of exchanges among citizens would represent a threat for civil liberties and would be considered as unconstitutional and not acceptable in many countries. In addition, the transterritoriality of the network would lead a foreign government to supervise information exchanges by individuals or organizations that do not act under their jurisdiction. Again it would be considered as unacceptable by many.
- More generally, private information spaces can be impenetrable for traditional enforcement authorities. Thanks to the ability to code information and to manage information spaces on a decentralized basis, individuals or groups can close information spaces. Moreover, these third parties would hardly identify anyone responsible in the last resort of potential legal rule-breaking. Indeed, the way information is managed in these information spaces can be faked, and the various operations can be distributed throughout the whole network, opening escape doors to the infringers who would easily be able to relocate their activities in the event of lawsuits in a particular jurisdiction area, being able at the same time to continue their activities. The users of Napsters switched to new systems like KaZaA and Gnutella, when major music companies sued the too centralized former system.

In order to have national laws enforced on the Internet, national governments should create a 'national Internet' with clear boundaries and the ability to control the exchanges (through gateways) with other national Internet

systems. Such architecture would, however, result in wide losses of positive network externalities, since it would *de facto* result in a bounded interconnectivity. Moreover, it would necessitate both the ability to effectively forbid any uncontrolled interconnection with a foreign network, and the ability to really control the exchanges of information of citizens and organizations with foreign counterparts. This would be costly, hard to legitimate, and would probably lead many users or potential users to switch to alternative information infrastructures. In other words, while it is technically possible to organize digital networks under the traditional hierarchical model controlled by national governments, such architecture would result in efficiency losses. First, users would no longer benefit from the generalized connectivity that enables them to finely organize information spaces according to their needs and preferences, without any territorial restrictions. Second, this decreasing quality of service would prevent the Internet from becoming a single and unified information infrastructure. It would maintain the existing information gaps due to the coexistence of heterogeneous information infrastructures, and would decrease the share of information activities benefiting from the efficiency gains brought by the use of digital technologies (agents being less incited to digitize their information activities, since the absence of a universal network would reduce the scale and scope effects of digitization).

Traditional institutional frameworks are also challenged in the digital economy because their logic and their legitimacy can be questioned. Of course, the legitimacy of organizing the delimitation and the allocation of uses over information on a territorial basis when a global network is available comes first. Before the rise of digital networks, it was relevant to do this since information flows were more intense within national boundaries than between them, and because the government was the only entity actually able to enforce exclusive rights of use. In a global information society, the legitimacy of such an organization is less convincing. Information goods can easily circulate worldwide, and a consistent system of right of use would increase efficiency by reducing the transaction costs over digital contents. Indeed the coexisting IPR systems force rights holders to claim for exclusiveness of use in many jurisdictions, to potentially sue infringers in these different jurisdictions and to manage complex contracts when transferring the right of uses to business partners or users that could manage operations in different countries. Redundancy is costly, but the cost of managing various IPR systems can become even higher when there are discrepancies among national laws.

The Internet is not only a-territorial (or transterritorial), it is also the heart of a global information infrastructure that supports the exchanges, the processing and the storage of all information flows, whatever their nature (voice,

image, text, data) and content. In the past, information networks were implemented on different and largely incompatible infrastructures, which were dedicated to specialized uses and which were made available only to users belonging to a same pre-existing community: an industry, a category of customers, companies of a certain type and so on. In many cases, specific regulations – a regulation being a way to implement or bound rights of uses – were implemented for each of these specialized networks because they were characterized by contrasted economics, technical capabilities and purpose. For instance, broadcasting licences were publicly granted because a scarce resource – the radio spectrum – had to be allocated in some way among conflicting uses and different operators. The restrictive regulations about contents in the public broadcasting of audio and video programmes (much tougher than for printed material) were justified by the technical difficulty of screening the various categories of audiences in the mass-media networks. Many of the justifications for the limitations of the rights to produce or communicate information over information networks are weakened or removed by the development of the Internet. Moreover, many of these past regulations of contents (and the various categories of intellectual property rights) can implement conflicting principles, which are impossible to manage in a unified information space. For instance, privacy is traditionally strongly protected in telecommunications networks. The content of an exchange, and even the existence of an exchange of information between two correspondents, cannot be screened and tracked by any third party, unless the judicial system provides authorization to document a case. On the other hand, the notion of privacy is meaningless in a broadcasting network. In the case of the Internet, what principle should be applied to information flows? Should a government or any entity be authorized to screen and track information exchanges? Should this principle be applied only to one-to-many communication? In the last case, what is the threshold? Is it really possible to identify the hidden one-to-many exchanges and so on? Clearly, many regulations of the past are no longer relevant.[13]

Moreover, the new technological context impacts on the efficiency of the former rules. For instance, copying copyrighted material for private purposes has been authorized in most IPR laws. This *de facto* restriction on the rights of IPR owners to control the use of their material was justified because copying did not really harm their capabilities of getting revenues. Private copy was limited in scale due to the cost and the low quality of copies. Today, the ability of digital technologies to make for free perfect copies that can be distributed on a very large scale could cut most of the revenues of copyright owners. If we admit that they have to get a return on their investments to create the intangibles, then this formerly justified restriction of copyright turns out to be totally inefficient.

Thus, traditional institutional frameworks in charge of organizing the use and the circulation of digital goods are challenged by the rise of global digital networks. On the one hand, digital technologies provide the producers and the users with tools to cheaply implement rules about their uses. On the other hand, the rules designed by traditional frameworks become less relevant than before (and can even generate major inefficiencies), while their enforceability is decreasing.

Many discussions about the required institutional frameworks to organize the cyber-world are fuzzy since there are two different sets of problems raised by the development of the global digital information infrastructure. First, the Internet is in itself a new economic space – a new frontier – in which rights of uses over resources are not yet totally and clearly established. A process has to be run to establish how rights of uses have to be delimited and allocated. Second, the Internet is the infrastructure on which a specific category of goods – information goods – is going to be produced and exchanged. When it comes to IPRs, we are considering intangibles only, we should therefore essentially focus on the second problem: the optimal design of a PR system for non-rival goods. However, while information goods are generally considered as non-rival, there are rival information goods. We shall start by discussing the case of this latter category of resources that are created in the new economics space before considering non-rival information goods.

Cyber-world: a new frontier

In the case of rival goods, the analysis of the optimal organization of a PR system on the Internet is close to the analysis of the optimal organization of a PR system in general, as developed by Barzel (1989), North (1990) and many others since most economists focus on the analysis of rival goods. There are in fact two main categories of rival resources on the Internet: addresses and signals of quality.[14] Before discussing how the management of these resources could and should be organized by the institutional framework, it is useful to return to some technical aspects of the Internet.

Addresses and signals as rival information goods

As stated above, the Internet is not a network *per se* but a set of principles and standards that enable any information processing device connected to a network implementing these principles and standards to be able to communicate and interoperate with any other IPDs connected to other networks relying on the same principles and standards. In addition to common, standardized interface languages, the performance of the resulting virtually unified network relies on a single 'addressing system', which allows any IPD to identify the other IPDs necessary to route the requests and the replies from the right client to the right server (see note 9), and vice versa.

On the Internet, the addressing system comprises two layers. First, a numerical address is allocated to each of the IPDs connected to the network: the the Internet protocol (IP) number. IP numbers are machine-only readable addresses that are the basis of the dialogue among the devices connected to the network. It is essential to avoid any duplication of IP addresses within the Internet, because it would prevent the clients from identifying the servers, and more generally disturb the routing of information among machines. Second, a 'user-friendly' addressing system – the domain name system (DNS) – is implemented to allow Internet users to express their request in a language that is close to 'human' language. The prefixes of the form 'www.identifier.com' are indeed easier to manage than IP numbers for bounded rational human beings. Moreover, this is a flexible system since the manager of a domain name (DN) can dedicate several IPDs (and therefore IP numbers) to a single DN. The nucleus of the DNS is a root file that establishes a single link between any DN and IP numbers. This allows any computer connected to the Internet to interpret requests expressed in HTML language (see note 8).

In fact IP addresses and DNs are different resources. An IP address can be considered as a mandatory registration to be included in the Internet system. Without IP, an IPD cannot operate on the Internet. The other machines connected to the network simply do not recognize it. A DN is not a mandatory resource to get access to the Internet as a consumer of contents (a 'client' in technical terms). For producers of contents, however, it is a means to facilitate access to their services. DNs free users to identify the IP numbers of the machines where information goods and services are localized. By decreasing considerably the search costs associated with the localization of contents, DNs are closer to signals of quality such as brand names, logos and labels than to addresses.[15]

IP numbers and DNs are rival resources since two different users cannot use them at the same time. Common IP numbers will simply lead to forbidding access to the Internet to the second party attempting to log on to the network. It could also hinder the performance of the network, resulting in a poorer service for the other users participating in the network. DNs are rival since if a party invests in the development of various capabilities to guarantee a level of service to its customers, and if this party invests in addition in communication to establish a direct link in the mind of the public between a symbol and this guaranteed level of quality, then the use of the same (or even a similar) signal by another player who would not guarantee its customers the same level of quality, will destroy the credibility, and therefore the usefulness and the value of the signal. Consumers will no longer benefit from the economies of search and inspection costs provided by a credible signal. Providers will lose the value of the investment they made in building a reputation. This might result in lower incentives to provide (both horizontally

and vertically) differentiated supply, while differentiation is an efficient reply to consumers' heterogeneous preferences.

These resources are not only rival, they are even scarce. Because of the required standardization and hierarchization of the system used to identify each of the IPDs connected to the network, there are a limited number of roots to create IP addresses. This causes a problem of allocation. One often quoted example is the University of Stanford which has the capacity to create more IP numbers than the People's Republic of China, because when the current addressing system was created the former had the opportunity to reserve large numbers of IP prefixes. With the implementation of the Internet or third generation, a new addressing system will become available (IP v 6). This should reduce this scarcity problem. However, the actual source of scarcity is in the DNS. The number of available names and expressions of the natural language that can be the base of meaningful addresses is obviously bounded.

Centralization as a guarantee for decentralized network operations

Thus the Internet *per se* is a new economic space where at least two categories of rival and intangible resources have to be managed: IP addresses and DNs. In each of these cases exclusive rights of use have to be delimited and allocated to users, both to simply enable the system to perform and to allow trade among them, should the initial distribution of these rights be enhanced to better fit agents' preferences. There are in each of these cases two extreme ways to organize the delineation and allocation of these exclusive rights of use. The decentralized solution is when the final users or the decentralized network operators – for example, the Internet service providers (ISPs), whether they are providing access on a commercial basis or not – self-claim exclusive rights of use. The centralized solution is when an authority of last resort is endowed with the right to make sovereign decisions in granting exclusive rights of use to claimants.

The advantages of centralization and decentralization will be discussed below. Before that, it has to be pointed out that any claim for exclusive rights of use is not self-enforceable in the specific case of IP addresses and DNs. In a fully decentralized system, and technically the Internet is fully decentralized,[16] anybody can claim for exclusivity on addresses, while nobody can force the other participants to recognize these exclusive rights of use. In particular, a new entrant could decide to use an already used IP or DN. In a fully decentralized system, the initial claim will not systematically be enforced, and the new claim might either disrupt the system or supplant the initial claim. A central register has therefore to establish the list of all the IP addresses to be recognized by the IPDs connected to the network, and to establish a one-to-one relationship between any DN and the related IP numbers. This central register must be

acknowledged (and its contents enforced) by all the users of the system.[17] Since self-enforcement of claims cannot occur, a minimum level of centralization is thus needed to ensure the absence of conflicting claims.

We shall now discuss more generally the advantages of centralization and decentralization, the impossibility of fully decentralizing the design of the addressing system being taken into account.

In the case of IP addresses, a decentralization of the concrete allocation of addresses to the users by Internet Service Providers (ISPs) – those entities providing access to the Internet – is the simplest way to concretely manage the allocation process. However, two very different decentralizations could be implemented. On the one hand, the ISPs can be endowed with an ability to distribute a pre-established list of addresses by the entity in charge of managing in the last resort the register of network addresses. On the other hand, the ISP can be allowed to develop its own addressing system to distribute as many addresses as it wants. It is therefore responsible for establishing a gateway between its own addressing system and other ISP addressing systems. The former solution does not allow for removing the intrinsic scarcity of available addresses. Moreover, it reduces the competition among ISPs since the number of granted addresses bounds their market share. This bounded competition could result in a lower quality of service and an inefficient allocation of addresses. The development of independent addressing systems in each subnetwork associated with the implementation of network address translator (NAT) resembles the current practice in traditional communication networks like the telephone system. It would solve the scarcity problem but would strongly decrease the transparency and the reliability of the network (because the addressing system would be composed of various layers). In addition, this solution would give a wide power of control to ISPs because the network would no longer be an end-to-end network.[18] ISPs would manage gateways between their networks and the other networks and would become therefore able to control what their users are doing. That could raise problems for two reasons.

First, it would allow ISPs to become private norm settlers, while they would not have to enforce any basic constitutional principles guaranteeing the protection of some fundamental rights to the users of the Internet – privacy, protection against arbitrary decision to remove rights of use, guarantees that in the case of infringement of its own rights of use, the last resort authority will ensure their enforcement and so on – since they are operating on a global market where such rights do not exist (Lemley 1999; Shiff Berman 2000) and since the competitive pressure can be weak (see Box 20.1). Moreover, for the same reasons, these private norm settlers can ignore the interests of the non-users of the Internet, while there are externalities,[19] resulting in potential capture of welfare to the detriment of these non-users.

Second, digital and network activities are characterized by the combination of fixed costs and increasing returns of adoption that make monopolies sustainable (see Box 20.1). The private norm settlers might therefore be freed from taking into account their customer's preferences when establishing their own norms. Moreover, their control over the addressing system would be a tool to establish market power since, with the collapse of end to end, network operators would become able to control the information service provision on their networks, and therefore to adopt strategies aimed at decreasing the competitive advantages of their competitors (in particular, by providing exclusive services on their own network). Not only would such strategies lead in the long run to the emergence of uncontestable monopolies, but they would also lead in the short term to a decreasing ability of Information Service Providers to market their services on the global digital market (with unavoidable consequences on the diversity and on the price benefiting the final users, since providers will have to write off the fixed cost of the service provision on a reduced audience).

For these reasons, the former solution is intrinsically superior to the latter: it preserves the end-to-end character of the network that is at the heart of its reliability and flexibility. It is, however, sustainable in the long run if and only if the IPV6 numbering plan can be implemented. In that case, the spectre of an actual scarcity of IP addresses would be removed, and the long-term sustainability of competition among ISPs guaranteed.

BOX 20.1 THE LONG-TERM STABILITY OF MONOPOLIES ON DIGITAL NETWORKS

The digital network economy is often considered an economy in which competition is sustainable because the decentralized nature of digital networks and the low level of barriers to entry seem to enable any victim of the exercise of monopoly power to bypass its service provider. In other words, contestability (Baumol et al. 1982) is supposed to be strong. Several scholars contest this oversimplistic conventional wisdom and point out that network or information service providers have some room for manoeuvre to create and exploit bottlenecks. For instance Crémer et al. (1999) or Tirole et al. (2001) emphasize that Internet operators can strategically decrease the transportation capacity of the network. By downgrading the quality of the interconnection with smaller networks, large network operators increase the relative quality of the services provided to their subscribers (whether they are final users or information providers) as compared to the service delivered by small networks.[1] Those who operate larger

networks are therefore able to attract the subscribers of smaller networks and to initiate concentration. Similar strategies can be observed on the market for content (Frischmann 2001; Posner 2000). Websites that benefit from the largest audiences are induced to develop various strategies to reduce the audience of the less-well-known ISPs and to expel them from the market. For instance, they can refuse to implement html links with the sites of their competitors. They can also sign exclusivity agreements with information or network service providers. Since positive network externalities arise, this type of ostracist strategy decreases the attractiveness of competing sites and reduces their visibility.

Such strategies can be harmful to the competitive process because barriers to entry exist. The required investments to develop broadband networks, for instance, or the communication costs to establish a new brand are significant (and these markets are already quite concentrated[2]). Due to the combination of increasing returns and positive network externalities – which are characteristic of information activities – incumbents benefit from strong protection once their market share is established.

The long-term viability and intensity of competition is therefore an essential challenge in the digital economy characterized by strong trends towards the emergence of viable monopolies (see Shapiro and Varian 1999; Noe and Parker 2000).

Notes

1. Indeed, subscribers of the 'small' networks have a larger probability than those of the 'big' networks of sending requests (request to access content or request to send information to a correspondent) to users that are reachable through a network that is not the same as the one they subscribe to. If interconnection is of poor quality, the service they receive is deteriorated (denial of access, long delays and so on).
2. Nearly 80 per cent of the Web traffic is dedicated to 0.5 per cent of the sites. The seven most important websites group around 20 per cent of the whole Web-supported data flows. The ISP market is also quite concentrated (see Gaudeul and Julien 2001).

In the case of DNs, decentralization is also the only way to concretely allocate addresses to users. The entity in charge of maintaining the root file of the DNS (that is, the source that establishes a link between each individual DN and IP numbers, and which is managed by ICANN in the present system) has to delegate to other entities (the registrars in the present system) the right to grant DNs to users, the essential problem being avoiding the allocation of the same DN to two users. However, in the case of DNs the problem is more complex than in the case of IP numbers. Indeed, DNs are not neutral since

they are used by Information Service Providers (whoever they are and whatever their motivations are) as signals. DN 'owners' want therefore to avoid confusion. In particular, if two DNs differ only by one or two letters, or only by a suffix, externalities could occur between the two owners. Moreover, DN claimants can consider that names and expressions are not insignificant. Meaningful expressions are scarce. Names with a specific reputation are also scarce. Lastly, exclusive rights of use can already be granted in the non-digital world (brand names, but also denomination of origin, family names and so on). Three types of conflict can then occur. First, claimants can have conflicting claims both because they want to be the single user of an exclusive DN and because they want to exclude other claimants from similar names. Second, registrars could also compete by distributing the more 'valuable' names. Third, the individuals and groups benefiting from the exclusive use of names outside the Internet can be harmed by the fact that alternative claimants would capture this exclusivity within the digital world, and would potentially either capture the investments made by the former to build reputation, or even decrease the value of this reputation. The problem is even more complex as the pre-existing rights of use of names can be conflicting since they were recognized in fragmented spaces (within national boundaries and often in even more local communities) resulting in the existence of several legitimate users of the same name.[20] Moreover, claims might target the private capture of names – such as names of celebrities, locations, events – and expressions that are considered as not individually appropriable (and therefore are common resources) outside of the Internet. In addition, the rules that apply to the usage of names can differ among jurisdictions.

The delineation and allocation of property rights over the DNS leads to the same problems as those raised in general when implementing a property system over a free but rival resource. Dividing the generic right to use a resource freely in a set of exclusive rights of use and distributing them among agents has a direct influence over the distribution of wealth among them. A decentralized process of negotiation – that is, a negotiation without any last resort arbitrator – can therefore hardly result in an agreement on the way the right to use the resource has to be divided and distributed, even if setting exclusive rights of use will increase collective efficiency (Libecap 2002). This is the well-known problem of collective choice raised by Condorcet (1785) and later by Arrow (1951). Note, moreover, that such an agreement not only covers the distribution of rights. Agents have to agree on the delimitation of these rights as well. For instance, should a right to use a DN cover a single suffix (like .com or .fr), or should it cover all the suffixes? Should the exclusive rights be permanent or temporary? Should these rights be extended to keywords? How do these rights overlap with other exclusive rights of using names such as brand names? Endless conflicts could result from such a

negotiation and agreeing on a conflict settlement procedure of last resort is essential to maintain the consistency and the operability of the DNS.

As in the case of IP addresses, a decentralized system of name distribution could be organized with different registrars (either on a commercial or on a geopolitical basis; see note 20) distributing freely their own DNs, and organizing, or not, gateways among their networks. This would result, however, in the coexistence of various Internets, since end to end and general connectivity would no longer be maintained. The various DNSs would *de facto* create independent networks. While this fragmentation of the network would be in that case weaker than if it were based on the management of alternative IP addressing systems, the fragmentation would be effective for most of the users and network externalities would be weaker. Moreover, because they will be granted a *de facto* exclusion right, the registrars would be able to implement their private norms, without caring about any essential constitutional principles protecting the individual users (as would be the case for network operators if the allocation of IP addresses were fully decentralized). For all these reasons, it would be costly to not implement a system guaranteeing in the last resort the consistency and the unity of the DNS, and more generally of the addressing system at the heart of the Internet.

Arbitrating among conflicting interests

Historically, primitive systems of property rights developed in a decentralized manner. However, it was in a logic of capture and pre-emption that in no way guaranteed either efficiency, or peace. As pointed out by North (1990), economic history shows that path dependency and rent seeking can prevent economic systems from evolving towards more efficient PR systems. In that context, the state often played a central role in arbitrating among the various interests under its jurisdiction – which does not mean that it is fair, benevolent and efficient – to help to implement PRs allowing either the reduction of transaction costs or the sustaining of growth. This is consistent with Libecap (2002) who suggests that a central entity might be useful in enabling the players to solve the redistribution problem – by implementing compensation – which hinders the privatization of a formerly free resource.

The procedures used to set up PRs in the present Internet in no way guarantee that the interest of all the stakeholders will be respected. Most of the existing norms result from a 'first-come, first-served' process. Inventors and early adopters were able to capture most resources. However, since the Internet is becoming the basis of many social interactions involving a wide range of different types of agent, there is no legitimacy to systematically adopting and enforcing the norms that have been designed by the first entrants, the stronger players, or the best-organized lobbies (Lemley 1999; Shiff Berman 2000).

Moreover, since there are interdependencies between the cyber-world and the actual one, even if norms were to result from a consensus ratified by the community of cyber-citizens, nothing would guarantee their efficiency, in the sense that the interest of every stakeholder would be taken into account (Lemley 1999). An efficient and fair distribution of resources requires processes and instances able to manage these externalities between the cyber and the real worlds.

To sum up, while the Internet allows a decentralized creation of PR, several elements call for some centralization in so far as it concerns rival resources. First, the addressing system of the Internet cannot be decentrally enforced, a last resort authority responsible for maintaining a central register of claims of exclusive use of addresses has to be managed to maintain the consistency and the unity of the addressing system that is at the heart of the end-to-end character of the network. Second, the measurement – that is, the delimitation and the allocation – of individual exclusive rights of use can be technically decentralized, but a central authority has to be in charge of granting these rights in the last resort. This is the only way to solve potential endless conflicting and overlapping claims, to avoid the infringements of exclusive rights granted outside the Internet, and above all to guarantee the end-to-end architecture of the network whose collapse could result both in reduced network externalities due to the fragmentation of the network and in monopolistic capture (without any ability to really limit because of the sustainability of monopolies in the digital economy).

The latter set of problems does not systematically occur for each of the two essential rival resources that are specific to the Internet (see Table 20.1). However, the combination of enforcement and measurement problems calls for some centralization and hierarchy within the Internet. The management of a decentralized network of networks implies an entity that sets the fundamental rights of the users, and that therefore bounds the ability of potential

Table 20.1 Potential problems raised by full decentralization of PR settlement

	IP addresses	DNs
Inconsistencies (overlapping claims/ unreachable agreement)	X	X
Infringements of existing rights		X
Private norm settlement without constitutional guarantees	X	X
Risk of monopoly capture	X	

private norm settlers to implement these norms without any limit. In addition, this entity of last resort should avoid inconsistencies in the delimitation and the allocation of these rights. However, given the size of the system and its complexity, it would be inefficient to manage it centrally. It is therefore essential to combine a decentralized management of the PR system with a principle of hierarchy and authority of last resort responsible for guaranteeing the unity of the system and at the same time some essential right of its users.

IPRs and non-rival goods on the Internet

The necessary decentralization of an IPR system

While some specific categories of information are rival, most of information and knowledge is of a public good nature. First, its consumption is indivisible. It is therefore worthy to have it as widely diffused as possible, while it is costly to produce it. This results in the well-known protection/diffusion dilemma. Second, it is 'naturally' complex (and costly) to exclude a third party from the use of information or knowledge since once revealed it is quite impossible to (voluntarily) remove it from the brain which stored it, and since reproduction costs, which are much lower than production costs, strongly decrease with the development of reproduction technologies, making it easy to multiply copies, whose circulation is complex to control.

Due to the latter characteristic, only agents who can physically constrain individuals can really control the circulation and use of information. Due to the former, it is generally acknowledged that the access and use of intangibles will be restricted in some way to enable their creators and potential investors to obtain a return on their investments and efforts, while it is recognized that this protection should be limited so as to favour diffusion ultimately (Besen and Raskind 1991).

This partly explains the organization of the traditional IPR systems (Bessy and Brousseau 1997, 1998). On the one hand, the intervention of the state is required, since its power of constraint in the last resort is necessary to prevent individuals from using information or knowledge even if they could easily access it. It is recognized at the same time that too strong a protection by the state would be inefficient since it would prevent diffusion. The traditional central (governmental) systems that 'measure' and enforce IPRs are therefore very incomplete, in the sense that they leave to the owners the responsibility for actually delineating and enforcing their rights of exclusive use. The producers of information or knowledge have to delineate the piece of information or the idea on which they claim exclusivity. The government only maintains a list (*cadastre*) where it registers claims for exclusiveness of usage. IPR owners are then responsible for detecting possible infringement and bringing

infringers to court to ensure that illegal use is stopped or damages and royalties are paid. Since IPR owners bear the cost of having their rights registered and enforced, the production–diffusion dilemma is partly solved, since only the more valuable uses of information goods are actually privatized by their creators, and because the search for exclusiveness is limited by the costs of claiming and maintaining exclusivity.

From a collective point of view, this is also a way of keeping the costs of the PR system at a reasonable level. Indeed, information and knowledge can be consumed in very different ways, and can be used as inputs to produce new information goods or services. Consider a piece of recorded music, for instance. It can be listened to by an individual, broadcast to a wide audience, played in a public space, used in a motion picture, be an input into a new piece of music and so on. Centrally measuring and enforcing exclusive rights for each of the possible uses of a piece of music would mean that a central public agency – unable to value the net outcome of protection for each possible use of each piece – would manage a cadastral database (registering all the rights' owners of each of these potential uses), and would check any audio-visual production and any use of music to assess whether exclusive rights of uses are infringed. Since defence of exclusive rights of uses is costly, agents *de facto* bound their claims for exclusiveness when protection is decentralized. This results in a less complete, but less costly, IPR system.

These pre-existing justifications for a substantial decentralization of IPR institutions is reinforced by the capabilities of self-regulation provided by digital technologies.

A central agency to ensure enforcement and prevent capture

While strong arguments call for a decentralized settlement of IPRs in the digital world, other arguments lead to the mitigation of this initial view. Three questions have to be addressed. First, whether self-claim and decentralized negotiation among claimants of exclusive rights of use would lead to inconsistencies. Second, would the self-enforcement of IPRs lead to a consistent and workable institutional environment? Third, would a decentralized IPR system result in delineation and allocation of rights of uses guaranteeing the best possible collective outcome?

First, as mentioned in the case of rival goods, in an economic space in which there are no *ex ante* legitimate exclusive rights of uses, a full decentralization of the delineation and claims of exclusive rights of uses could result in overlaps (conflicting claims on the same resource) and inconsistencies (no claim for some exclusive rights of use of some resources, resulting *ex post* in potential conflicts or lack of investment to develop and maintain these free resources).

Would that be a problem for non-rival goods? In fact, everything depends upon the centralization of the enforcement mechanism. If it is centralized –

that is, if an authority[21] recognizes and makes enforceable the claims for exclusive use – then exclusion actually applies and one comes back to the reasoning given above about rival goods. If there is no such central authority to enforce these claims, and agents therefore spend resources locally to exclude others (for example, by encrypting their contents), then conflicting claims are not an issue, since they result in competition on the supply of twice-claimed exclusive rights of use. Since information is not a rival good, claims for exclusive uses compete but do not conflict. The decentralization of claims – of the 'measurement' of IPRs – is therefore neutral as long as the enforcement is decentralized.[22]

Second, is the complete decentralization of IPR enforcement sustainable? As mentioned above, the enforcement of PRs relies generally on owners' efforts together with the intervention of a last resort authority that benefits from lower costs and extended ability and credibility in punishing infringers. At least two reasons justify the intervention of such a mechanism of enforcement in the last resort. First, no cryptographic system is inviolable. Code-based protection is therefore imperfect.[23] Second, the enforceability of collective norms founding virtual communities is also in question. The self-enforceability of norms governing virtual communities is bounded by the ability of Internet users to access alternative communities, providing them with the same type of service, and by the ability of the supervision mechanisms of these communities to identify rule breakers and to actually expel them from the community. Indeed, Internet users can hide their behaviours, and the only identity that is certain over the Internet is that of each computer. Put another way, a code cannot guarantee *ex ante* the enforcement of rules (in general, including therefore the rules concerning the rights to use information sequences) and communities are bounded in their ability to supervise behaviours and retaliate in the case of rule infringement.[24]

These call for a central mechanism granted with the power of constraint aimed at guaranteeing a minimum transparency so as to control how the protected contents are actually used. Indeed, virtual communities could be organized to hide the sharing of decrypted digital contents without the consent of owners. A mechanism limiting faking capacities would enable the owners of exclusive rights to enhance their supervision capabilities. In addition, and more fundamentally, there is a need for some exercise of constraint in the last resort. Indeed even if a self- or local regulatory mechanism can rely on capabilities to harm infringers, its credibility is bounded by the cost borne by the infringers in the case of exclusion (and by the costs borne by the entity exercising retaliations). If this cost is inferior to the benefit the infringers can make by violating the local regulation, the self-enforcement mechanism will be unable to prevent (major) infringements. A last resort enforcement device, which would be able to increase the costs of violating

the local regulations by implementing additional retaliation, would reinforce the enforceability of local regulation (see Milgrom et al. 1990; Brousseau and Fares 2000).

Behind the argument that a complete system of PRs could be entirely decentrally established thanks to information and communication technologies (ICTs), there is the assumption that PR settlements would cost nothing with the use of the technology. This assumption, which is implicit, could also lead to the opposite conclusion that a complete IPR system could be centrally created. In fact, lower costs of PR settlement do not mean that these costs are zero.[25] Encrypting digital information does not result therefore in a perfect control over its uses. Additional means have to be used to complete the incompleteness of these PRs. A central (traditional) PR institutional system could for instance back up the technological self-implemented protection (as in the music industry today). Economic agents could also try to decentrally reinforce technological locks by settling interindividual agreements. However, these contracts would necessitate an authority of last resort to guarantee their enforcement. Some centralization would therefore be necessary anyway. Positive measurement and enforcement costs unavoidably involves agents in having to mix centralization and decentralization to reduce these costs.

Third, beyond the question of the capability of a process based on self-claim and self-enforcement to generate consistent and workable PRs, the ability of such a decentralized process to generate efficiency has to be questioned, as well. If we admit that alternative institutional frameworks have contrasting impacts in terms of collective efficiency and distribution of wealth, some mechanisms to aggregate individual preferences have to be designed to select (even arbitrarily and imperfectly) a collective regulatory framework that would seek to result in the best collective outcome. Indeed, if one admits that various stakeholders have various interests and do not have the same ability to influence the decentralized settlement of IPRs, either because they have contrasting endowment or because they enter in the game at different periods while there are path-dependent phenomena (such as pre-emption), then decentralization does not guarantee that the delineation and allocation of PRs would be fair – by considering individual interests – and efficient – by allocating rights of use so as to reach the best possible use of resources from a collective point of view. If one admits that stakeholders can have conflicting interests, one should admit that some type of centralization is needed to compare the various possible PR systems in terms of collective efficiency so as to choose the most efficient one, whatever the efficiency criteria is. The important point here is to design a process in which the interests of the wider set of stakeholders would be taken into account.

In the specific case of information and knowledge, a PR system based on self-claims and self-enforcement could in particular lead to excessive and

indefinite private capture of these public goods. It would lead to distortive capture of monopoly rents, and to the hindering of the efficient diffusion of that information. Indeed, while information becomes a good whose uses are now eligible for technical exclusion, it remains a non-rival and indivisible good. It is therefore legitimate to question the optimal level of protection within the traditional debate that balances the advantages of strong incentives to create intangibles with those of a wide diffusion (see Besen and Raskind 1991). The example of freeware shows that sharing information on a very large scale maximizes the benefits of disclosure. In some cases, mandatory disclosure rules – especially if disclosure rules could be tailored to different audiences; for example, free access to published material to students and teachers in low-income countries – would be collectively optimal. Such rules could spontaneously emerge, as happened in the open-source software communities. However, there are also many situations in which it is doubtful that they would. Since investors in the creation of digital sequences could fear they would not get any return if they disclosed them, and since they can cheaply control access to them, it should result in an IPR system that would overprotect resources; in the sense that the collectivity would be deprived from the potential positive externalities of open access (externalities of diffusion, spillovers and so on). This calls for some centrality, both to select the collectively optimal system, and to compensate those who are harmed by a system in which PRs are bounded as compared to what they would get if a system of unlimited rights to control uses prevailed.

There is an additional reason for bounded PRs. The ability to use specific information or knowledge often depends upon the access to complementary information. A central system guaranteeing some fundamental rights of access to information should be able to guarantee the exploitation of these externalities.

The limitation of exclusive rights of use over information and knowledge should combine disclosure rules and the limitation of encryption capabilities. Indeed, the combination of the capability to encrypt with the long-term viability of monopolies leads to the possibility of controlling and preventing the free diffusion of information. This capability could be used in particular to reinforce barriers to entry, so as to allow endless capture of rents and the blocking of innovation and creation in the long run (by forbidding access and use of existing creations of the human mind). Such threats should lead to the bounding of agents' encryption capabilities (for example, mandatory registration of code keys to trustworthy third parties) so as to maintain a minimal level of transparency aimed at allowing supervision by some last resort authority in charge of preventing monopolistic capture. Moreover, reducing encryption capabilities would limit *de facto* the levels of barriers to entry, and therefore the strength of monopoly power.

The self-enforcement of PRs would therefore have to be supervised by some last resort authority ensuring that encryption and self-regulation are not combined to develop and exercise monopoly power, and to maintain competition in the long run. Indeed, competition is the best solution to provide agents and communities with incentives to implement efficient technical and organizational solutions in terms of knowledge production, use and sharing.

It has to be pointed out that bounding encryption capabilities would not be enough *per se* to ensure efficiency in the long run. Supervision of behaviour by a third party is essential since it would dissuade people from taking anticompetitive action, without forcing agents to be totally transparent about their behaviours and information exchanges. The protection of contents (both the privacy of information exchanges and property rights) leads to encryption, and it is not justified to broadcast publicly all information exchanges. It is, however, necessary to verify that information exchanges are not harmful for the collectivity as could be the case if they were aimed at setting up collusive agreements, infringing IPRs or performing criminal activities. An independent and neutral third party in charge of supervision is therefore a good solution to deter indefinite monopolistic capture, while enabling the agents to preserve privacy and to protect access to their informational contents.

To conclude, since the code and the lists of subscribers (to a virtual community) are not perfect enforcement tools, and since information remains a non-rival good, a decentralized IPR system on the Internet could lead to inefficiencies. On the one hand, agents may have to dedicate too many resources to the enforcement of their claimed exclusive rights of use. On the other hand, information could be overprotected, both because the optimal level of diffusion would not be reached and because some players would use information protection to establish monopoly power. A last resort authority, in charge of guaranteeing the enforcement of rules established by local regulators, while guaranteeing that these rules result in a desirable level of diffusion, and that encryption cannot be used to hide contents and behaviours endlessly, would therefore be useful to guarantee a more efficient outcome of a decentralized process of IPR settlement.

Organizing a PR system on the Internet

Digital technologies challenge the relative efficiency of the existing alternative institutional frameworks. However, the decentralized and unorganized process of production of self-legitimated norms does not guarantee that the resulting norms will be either workable, or efficient. All these call for the organization of an institutional framework that will enable these weaknesses to be overcome. Stakes are huge since the Internet, as an infrastructure, is becoming an essential facility on which economic activity takes place, on

which individuals communicate and share information and knowledge, and through which collective goods are provided to citizens.

The decentralized process of PR settlement based on self-claims and self-enforcement that is made possible by the use of digital technologies has to be combined with central coordination aimed at avoiding the inefficiencies and weaknesses of self-regulation, and at implementing the most efficient solutions by taking into account the interests of the widest possible set of stakeholders, the public nature of information and the risks of having digital technologies used to deter competition and capture rents without time, or scale, or scope limits. It would therefore be worthwhile to set up a last resort authority, which would have to design and make enforceable constitutional principles aimed at guaranteeing some fundamental rights both for contents producers and for users. While decentralized systems of IPR settlement would enable agents to benefit from coordination frameworks well adapted to their specific needs and preferences, the last resort institution would maintain the consistency of the resulting systems of private norms, would ensure the enforceability of these self-regulations,[26] and would prevent capture of public wealth by individuals or groups. This calls for a federal institutional model enforcing a subsidiarity principle. The central and last resort institution is there to guarantee the efficiency of a decentralized mode of self-regulation, not to directly regulate uses.

This last resort regulation device should be submitted to democratic control and be responsible for enforcing a basic constitution aimed at preventing capture and protecting essential natural rights. It should act more as a jurisdiction than as a government. However, it has to be made clear that, as a regulator it will both settle conflicts and design the constitutional rules.

Beyond its logical justification, the implementation of a regulator of last resort is made possible on the Internet by the necessity of managing the addressing system centrally. The mastering of the management of the addressing system by the entity that would be responsible for the regulation in the last resort will allow this entity to dispose of the means of its assignment. Indeed, it would enable it to dispose of a credible threat – excluding agents from access to the cyber-world by depriving them of IP addresses – that it could use to have its decisions and regulations respected. In turn, only a well-designed and democratically controlled entity should be allowed to control the system of inclusion in or expulsion from the Internet.

Notes

* This chapter was written when the author was a fellow at the International Centre for Economic Research (ICER, Turin), which is thanked for its support. It benefited from comments of participants in workshops held at the University of Turin, Statistic Norway, and the University of Paris I (ATOM). An earlier and broader version ('Internet regulation: does self-regulation require an institutional framework?') elicited useful suggestions

from Bruno Chaves (ATOM), Nicolas Curien (CNAM), Godefroy Dang N'Guyen (ENST-B), Mhand Fares (INRA), Didier Lebrun (ISOC), Hervé Nabarette (ATOM), Lee Davis (CBS) and Birgitte Andersen (Birkbeck, University of London). I warmly thank them all. Usual caveats apply.

1. This very fundamental understanding of property rights clearly differs from the legal notion of property rights, but the latter is included in the former. Barzel's and North's analyses aim at synthesizing how an institutional framework results in economic properties through the allocation of decision rights to agents and through their ability to be exchanged (to arrive at a more efficient use of resources). Their definition of property rights encompasses what we usually call property rights, but also the regulations that bound economic agents' decision rights, the law that establishes boundaries to the free will of agents, and the informal rules (such as social customs) that mitigate and frame individual freedom of decision, and contracts since they allow agents to delineate the rights of use transferred to other agents.
2. It can also be said in a more positive way: the central authority can ensure *ex ante* the absence of conflicting claims *ex post*, since it is able to design a system avoiding inconsistencies – mainly overlaps – among claimed rights of use.
3. When externalities occur, the price system does not reflect all the constraints of the economy. Going back to Coase (1960), we acknowledge and agree that 'externalities' are not natural characteristics, but the consequences of the incompleteness of the PR systems. However, if we agree that in any case, setting up a PR system is costly, then we must recognize that any PR system is incomplete (since the marginal pay-off of delimitating and enforcing exclusive rights of use falls beyond the marginal costs of doing so at some point). Whether it is set up centrally or decentrally, an incomplete PR system generates externalities.
4. Of course, since the producers of these resources should get a return on their investments, the central authority can decide either to subsidize them in exchange for producing free goods (as in the 'public science' model), or to grant them with temporary and bounded exclusive rights of uses (as in the patent-copyright system).
5. Monopoly capture is inefficient in particular when it generates an underoptimal use of available resources, or when it hinders the dynamics of investment and innovation. As soon as the use of a resource is in some way not totally rival, a profit-maximizing monopoly will deprive some of the potential users from access, resulting in inefficiencies (especially if discrimination is costly and therefore imperfect). When network effects (and in particular, increasing return of adoption) occur, a monopoly can block either the innovation process or the adoption of alternative solutions. Monopoly capture is then dynamically inefficient.
6. Completeness of property rights corresponds to an ideal. It would mean that any potential use of any resources would be identified and associated with a right granted to either an individual or a group. A complete PR system would enable almost costless transactions since agents would only have to bear the cost of meeting and agreeing on the terms of the exchange. Transaction would not be free, however, because the costs of the PR system would have to be taken into account.
7. Indeed, in a positive perspective one should admit that selection processes are not perfect, and that public institutions are subject to capture by private interests (whether they are bureaucrats, politicians or groups of interests). Both can prevent an efficient institutional design from emerging.
8. The Internet is based on the use of two types of standards. The Internet Protocol (IP) is the common communication protocol that makes it possible to manage data flows among information processing devices (IPDs). It is the heart of the interoperability of the components or the various networks involved. The hypertext markup language (HTML) is the multimedia language of the Internet that enables any IPD to transform any kind of information (data, sound, image and so on) into codes that can be 'understood' by any other IPD. This is a common programming language that allows heterogeneous devices to interoperate when processing information.
9. From a logical point of view, the Internet relies on two basic principles. First, each IPD

that is connected to the network plays two roles simultaneously: processing information and operating the network. In a digital network, there is no technical discrimination between the resources dedicated to the administration of the network and the terminals that process the information carried, as is usual in traditional communication (for example, telephone) or distribution (for example, TV broadcasting) networks. In the latter, the network operator is responsible for managing transportation and switching capabilities to ensure the exchange of information among 'terminals' that do not interfere in the administration of the network. In a digital network each IPD is a switch that receives information from the other IPDs and routes it to the targeted IPD. Even though, in practice, some IPDs are specialized in the management of data flows, each IPD connected to the Internet has some routing capacities. This is the key to the decentralized administration of the network. Second, all the services provided by the Internet rely on client–server architecture. Any IPD on the Internet can become a client that sends requests to another IPD – which then becomes a server – to provide the former with information processing or service. The combination of these two principles makes it possible to generate any communication or information services provided by the Internet, and the development of new services relies on the addition of new IPDs (or software) that enrich the collection of basic services that can be combined to produce the various available services.

10. This is typically what happens today in the music industry. Major music companies try to code the recorded music they sell to avoid the duplication of their digital files. Generally, for instance, they make recorded songs available on the Internet to enable potential customers to listen. However, these digital files are coded either to reduce the quality of copies or to prevent copies from being made. Especially for most famous artists, hackers attempt to break these coded protections to display the music for free. In response, music companies develop tools to harm hackers' information systems. For instance, in case of infringement, viruses that attack the hacker's system can be implemented in the code. Music companies are also suspected of employing former hackers to track the hackers and attack their systems.
11. Indeed, software literally 'embodies' knowledge and makes it operable, since it implements routines to solve all kinds of problems. However, knowledge is also 'embodied' in databases, in text or audio-files explaining theories or ways to operate and so on.
12. It is often claimed that this is only true for small players, because governments are unable to detect and bring to court all the small infringements made by individuals and small and medium-sized enterprises. Moreover, if these individuals and small businesses do not have any assets located in the country where they infringe the law, the local government cannot retaliate. On the other hand, large companies should comply with most existing legal constraints since their behaviour is 'visible', and since they have in any case some form of tangible anchoring in all the countries where they have operations. Moreover, governments could try to deter them from breaking the law by threatening to harm their commercial reputation. Such arguments are weakened since large firms could conceal information-handling processes, and it can be costly for governments to prove that the law was broken. Moreover even if only small infringers are able to break the law with *de facto* impunity, billions of them could overwhelm traditional frameworks. This is typically what is happening today in the music industry.
13. Because the present Internet is still an imperfect substitute for most traditional network services – telephony, radio broadcasting, TV and so on – existing regulations can be maintained because bypass possibilities are limited. However, the development of broadband Internet, and the rise of a wide set of complementary technologies – such as e-books or printing on demand – will transform digital networks into a unified support for the diffusion and use of any type of content that will inevitably turn former regulations into illegitimate ones.
14. The communication capacity – the bandwidth – is the third rival resource on the Internet. That said, the Internet relies on principles that make the competition for bandwidth less acute in the Internet than in any other network. The end-to-end connectivity makes the totality of the remaining bandwidth available for marginal communication. Despite this principle, however, bandwidth remains a scarce resource. At any moment in time, it is

bounded by the capacity of the infrastructure and by the capacity of the critical nodes of the network (whether they are interconnection points or servers). Two types of problem then arise. First, a criterion to allocate the available bandwidth at each period has to be established. Second, the Internet operators must be encouraged to invest to limit the risks of network congestion (Frischmann 2001). While bandwidth is clearly a rival resource and raises PR delineation and allocation problems, it cannot be considered as an IPR problem. However, the analysis developed hereafter could be applied to the optimal design of a system to manage rights of use over bandwidth.

15. That has been *de facto* recognized by the users of the Internet and by those in charge of the management of the addressing system since the end of the 1990s. ICANN (see note 17) – in close cooperation with the World Intellectual Property Organization (WIPO) – progressively set up processes to recognize brand names and trademarks within the DNS. While in the early days of the commercial Internet, any user was allowed to register for exclusiveness in using any type of identifier as a DN – leading to the emergence of cybersquatting, a practice consisting in capturing famous names in order to resell them later to those who had made them famous or in order to develop services aimed at capturing customers valuing the reputation associated to this name – procedures have progressively been defined to enable owners of brand names to benefit from priority in getting exclusive rights to use these names in the DNS.
16. In fact, this is true from a logical, but not from an operational point of view. The Internet could be fully decentralized, but it is simpler and more efficient to maintain some elements of centralization. Network servers are intermediaries among the IPDs connected to the network. They take charge (both on the client and server sides) of the management of communication. This ranking of the network simplifies its topography and facilitates the routing of information among IPDs, but it is not mandatory to rank an end-to-end network. The key element of centralization, however, is the root server of the DNS. The nucleus of the DNS is a root file that establishes a single link between any DN and IP numbers. This root file is duplicated in several root servers that are themselves used by the IPDs connected to the network to address their communication to the right IPD. Having a limited number of copies of the root file facilitates the maintenance of the system. However, the root file could logically be copied in any IPD connected to the network to suppress any kind of technical centralization.
17. ICANN (Internet Corporation for Assigned Names and Numbers; www.icann.org) is the organization currently in charge of 'governing' the addressing system of the Internet. It is responsible for distributing IP numbers and DNs. It is a non-profit organization incorporated in the United States and set up in 1998. It operates under a delegation contract with the US government (Department of Commerce).
18. See the Internet Transparency, RFC-2775: ftp://ftp.rfc-editor.org/in-notes/rfc2775.txt; Blumenthal and Clark (2001).
19. For instance, if systems that allow large-scale barter of private copies of digital contents (for example, Napster or Gnutella) develop, the revenues of the creators of content will be affected, unless taxpayers are asked to compensate by contributing more to the funding of the production of works of art. In both cases, it is clear that the norm of free exchange applied by the members of an on-line community affects the welfare of members of off-line communities.
20. In a sense, this problem has been being addressed with the development of the numbers of 'domains'. ICANN promoted the development of national domains (country code top level domains – ccTLDs – corresponding to the suffixes .fr, .uk, .cn and so on for each of the countries recognized by the United Nations) and of additional generic domains (global top level domains – GTLDs – corresponding to the suffixes .com, .org, .info, supposed to correspond to the type of organization that claim for exclusive uses rights [respectively, for the suffixes quoted above: commercial corporations, not-for-profit organizations, information providers]). However, there is clearly a conflict between the logic of the ccTLDs – which *de facto* endorses the existing geographical fragmentation of 'property rights' on names – and the logic of GTLDs – which tends to establish global property rights on names. Moreover, the 'reputation' of the various top level domains is unequal – .com being, for

instance, the symbol of the Internet because it is understood as 'communication' rather than as 'commercial' by the public – creating a *de facto* hierarchy among the TLDs. The 'game' is also complicated by the fact that some countries with a specific suffix try to turn their ccTLDs into GTLDs – like the Tuvalu Islands attempting to grant any television channel in the world a domain name with the suffix .tv – and that there are many attempts to multiply the number of TLDs in order to create new 'labellized' information spaces: for example, .eu to be applied to any type of organization within the European Union, or .kids to be applied to sites targeting children with controlled contents. The resulting fuzziness leads clearly to a *de facto* globalization of property rights on names, in particular of brand names. Many owners of DNs, tend to register the same name with all the possible (and available) suffixes, whether they are GTLDs or ccTLDs. Indeed, famous brands want to avoid the new form of cyber-squatting that could develop with the multiplication of suffixes.

21. The authority can be a government, an intergovernmental agency, or a business alliance that would be able to deny access to contents to unentitled third parties. Think, for instance, of an alliance among major music companies.
22. If it were not, it would become a problem. However, decentralized claims are inconsistent with centralized enforcement. Indeed, in the latter case, the central supervision of uses would imply at least a mechanism that will centrally register claims of exclusive rights of use, which would have to check for the legitimacy and the absence of overlap among these claims.
23. This is typically the reason why the United States passed the Digital Millennium Copyright Act (DMCA) in 1998 to protect the code that protected contents, by severely punishing techniques and practices aimed at cracking codes. However, the DMCA did not take into account the necessity to guarantee openness in exchange of stronger institutional protection. See the end of this section.
24. The literature on private norms often refers to historical experiences (such as the medieval 'Law of Merchants') or to those norms that regulate many ethnic communities to point out the efficiency of decentralized self-regulation (see Granovetter 1985; Bernstein 1992, 1996; Cooter 1994, 1996). However, some papers also point out the limitations of self-regulation (see Milgrom et al. 1990). We pursue here such types of analysis. Indeed, and as pointed out by Lemley (1999), when ostracism does not apply, norms have to be enforced by external coercion mechanisms that can exercise some power of last resort over those who are supposed to enforce these norms.
25. First, data processing costs are not zero and encrypting information is costly (for example, it takes time, generates failures and so on). Second, information-processing costs are often fallaciously confused with data processing costs. ICTs impact less on the first category than on the second. Indeed, to manage complex information management processes, it is necessary to benefit from the ability of the human brain to combine heterogeneous cognitive processes.
26. Indeed, since these local regulations can be considered as components that participate in the general efficiency provided by the institutional framework, it is legitimate to reinforce their enforceability when necessary. The authority responsible in the last resort for the regulation of the system has therefore to use its credible threats to punish infringers of self-regulations. This is a common practice in the real world when the state becomes the guarantor of the enforceability of self-elaborated norms, like professional codes of conduct, by making them legally binding (see Bessy and Brousseau 1998; Brousseau and Fares 2000).

References

Arrow, K.J. (1951), 'Alternative approaches to the theory of choice in risk-taking situations', *Econometrica*, **17**, 404–37.

Barzel, Yoram (1989), *Economic Analysis of Property Rights*, Cambridge: Cambridge University Press.

Baumol, William J., John C. Panzar and Robert D. Willig (1982), *Contestable Markets and the Theory of Industry Structure*, San Diego, CA: Harcourt Brace Jovanovich.

Bernstein, L. (1992), 'Opting out of the legal system: extra legal contractual relations in the diamond industry', *Journal of Legal Studies*, **21**, 115–57.
Bernstein, L. (1996), 'Merchant law in a merchant court: rethinking the code's search for immanent business norms', *University of Pennsylvania Law Review*, **144** (5), 1765–821.
Besen, S.M. and L.J. Raskind (1991), 'An introduction to the law and economics of intellectual property', *Journal of Economic Perspectives*, **5**, 3–27.
Bessy, Ch. and E. Brousseau (1997), 'The governance of intellectual property rights: patents and copyrights in France and in the US', Inaugural Conference for the International Society for New Institutional Economics, 'The Present and Future of the New Institutional Economics', 19–21 September, Washington University, St. Louis, MO.
Bessy, Ch. and E. Brousseau (1998), 'Technology licensing contracts: features and diversity', *International Review of Law and Economics*, **18**, 451–89.
Blumenthal, M.S. and D.D. Clark (2001), 'Rethinking the design of the Internet: the end-to-end arguments vs. the brave new world', in B. Compaine and S. Greenstein (eds), *Communications Policy in Transition: The Internet and Beyond*, Cambridge, MA: MIT Press, http://www.ana.lcs.mit.edu/anaweb/PDF/Rethinking_2001.pdf, September.
Brousseau, E. and M. Fares (2000), 'The incomplete contract theory and the new institutional economics approaches to contracts: substitutes or complements?', in C. Ménard (ed.), *Institutions, Contracts and Organizations: Perspectives from New Institutional Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 399–421.
Coase, R. (1960), 'The problem of social cost', *Journal of Law and Economics*, **3**, 1–42.
Condorcet, Marie Jean Antoine Nicolas Caritat (Marquis de) (1785), *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix* [Essay on the Application of Analysis to the Probability of Majority Decisions], Electronic edition (1995), http://gallica.bnf.fr/scripts/ConsultationTout.exe?E=0&O=N041718, accessed 2002.
Cooter, R.D. (1994), 'Structural adjudication and the new law merchant: a model of decentralized law', *International Review of Law and Economics*, **14**, 215–31.
Cooter, R.D. (1996), 'Decentralized law for a complex economy: the structural approach to adjudicating for the new law merchant', *University of Pennsylvania Law Review*, **144** (5), 1643–96.
Crémer, J., P. Rey and J. Tirole (1999), 'Connectivity in the commercial Internet', Working Paper, Institut d'Économie Industrielle (IDEI), Toulouse.
David, P.A. (1985), 'Clio and the economics of QWERTY', *American Economic Review*, **75** (2), May, 332–7.
Frischmann, B.M. (2001), 'Privatization and commercialization of the Internet: rethinking market intervention into government and government intervention into the market', *Columbia Science and Technology Law Review*, vol. II (2000–2001), http://www.stlr.org/html/archive/, accessed 2002.
Gaudeul, A. and B. Julien (2001), 'E-commerce, vers une analyse économique' [E-Commerce: Some Elements of Industrial Organization], *Revue Économique*, special issue, 'Économie de l'Internet', edited by E. Brousseau and N. Curien, **52** (HS), October, 97–118.
Granovetter, M. (1985), 'Economic action and social structure: the problem of embeddedness', *American Journal of Sociology*, **91** (3), November, 481–510.
Lemley, M.A. (1999), 'The law and economics of the internet norms', Working Paper, University of California at Berkeley.
Lessig, Lawrence (1999), *Code and other Laws of Cyberspace*, New York: Basic Books.
Libecap, G.D. (2002), 'A transactions costs approach to the analysis of property rights', in E. Brousseau and J.M. Glachant (eds), *The Economics of Contracts: Theories and Applications*, Cambridge: Cambridge University Press, pp. 140–56.
Milgrom, P., D.C. North and B. Weingast (1990), 'The role of institutions in the revival of trade: the law merchant, private judges, and the Champagne fairs', *Economics and Politics*, **2** (1), March, 1–23.
Noe, Th. H. and G. Parker (2000), 'Winner take all: competition, strategy, and the structure of returns in the Internet economy', Mimeo, Tulane University, November.
North, Douglass C. (1990), *Institutions, Institutional Change and Economic Performance*, Cambridge: Cambridge University Press.

Posner, R.A. (2000), 'Antitrust in the New Economy', John M. Olin Law & Economics Working Paper 106, Law School, University of Chicago, November, http://www.law.uchicago.edu/Publications/Working/, accessed 2001.
Shapiro, Carl and Hal R. Varian (1999), *Information and Rules*, Cambridge, MA: Harvard Business School Press.
Shiff Berman, P. (2000), 'Cyberspace and the state action debate: the cultural value of applying constitutional norms to "Private" Regulation', *University of Colorado Law Review*, **7** (4), Fall, 1263–310.
Tirole, J., J.-J. Laffont, S. Marcus and P. Rey (2001), 'Internet interconnection and the off-net-cost pricing principle', Working Paper, IDEI, Toulouse.

21 Collective property rights for economic development: the case of the ceramics cultural district in Caltagirone, Sicily

*Tiziana Cuccia and Walter Santagata**

Introduction

Murano is famous all over the world for its exquisite glassware, created by designers with great aesthetic taste and refined technological know-how. In Arezzo, Vicenza and Valenza Po, hundreds of goldsmiths work precious metals and diamonds using a combination of traditional techniques and original creativity: Como's printed silks, Prato's fabrics, Biella's fine wool, and Faenza's, Albissola's and Caltagirone's artistic ceramics are other examples of localized production of culture-based goods. This diffused system of localized small firms with a high aesthetic and intellectual content makes Italy the land of cultural districts par excellence (Santagata 2002a, 2004). Italy, more than other countries, provides favourable conditions for so many small firms to flourish, for community cultures to thrive; for the development of the necessary know-how and conditions to turn an area of craftsmanship into an industrial cultural district.

The aim of this chapter is to analyse the institutional aspects of the ceramics industry in the cultural district of Caltagirone. As the market of intangibles goods (ideas, forms and design) develops and expands, protecting intellectual property becomes a key issue. Therefore public policies supporting the sustainable economic development of localized industry become necessary (Benghozi and Santagata 2001). From this perspective, we believe that establishing individual and collective property rights would facilitate entry into the post-Fordist industrial phase of organizational flexibility and global markets.

This chapter proceeds as follows. First, Caltagirone's ceramics industry is described and explained as a cultural district, where a range of idiosyncratic localized culture-based goods are manufactured. Then we shall examine the major results of a survey carried out in this district (Cuccia et al. 2001). In particular, we shall describe the individual preference for the establishment of a local system of property rights. Furthermore, we shall make use of a game theory approach in order to explain the behaviour of Caltagirone's *cannatari* (ceramists) with reference to trademarks and the establishment of a collective system of property rights. Attention will be drawn to the advantages and disadvantages that the creation and manage-

ment of a collective distinctive mark might bring about in terms of social welfare.

Idiosyncrasy and universality of Caltagirone's ceramics as culture-based goods

To understand the social and entrepreneurial phenomena of a cultural district it is necessary to trace the origins and idiosyncrasies of that culture.

Cultural processes are idiosyncratic and universal. These two major characteristics are not contradictory, but are in fact complementary. Intangible culture has its roots in the geographic area where its identity and creative power emerge. At the same time, its message and evocative presence are everywhere; they are universal and speak every language. Idiosyncrasy refers to the roots of a culture, its ties with the local society, with the history of its ruling classes and institutions, and with the natural resources of the region. It proves that a specific production did not start by chance, and that all community members share that intangible culture and can become entrepreneurs. In contrast, universality is the intangible and boundless value of a cultural good, containing ideas, beauty and quality that can be understood worldwide. The consumption of culture-based goods and services is ubiquitous in time and space.

The concept of culture as an idiosyncratic, particular and localized good and its connection with the theory of cultural districts (Scott 2000; Santagata 2002a, 2004) can be further subdivided, and this will be illustrated using Caltagirone's ceramics as an example.

Social origins of a localized cultural production

Two essential factors underlie the origin of localized culture: first, the availability of natural resources which are utilized by the local inhabitants to develop a specialized industry that can be handed down to subsequent generations, and second, the emergence of one or more ruling social classes, who demand quality or luxury goods, and who interact with the external world for aristocratic, religious, military or commercial reasons.

According to this interpretation, the idea that culture and creativity are the unforeseeable and unexpected results of genius and talent is to be integrated with the fact that discoveries and innovations, including cultural ones, are not accidental and never occur by chance – they are always the result of an organized and well-structured research project. The cultural *genius loci* consists in natural resources and an emerging social class structure.

The production of Caltagirone's ceramics is the expression of a local tradition with Sikeliot origins. This is confirmed by the discovery of a large pot, dated 5 BC, illustrated with red motifs portraying a ceramist's atelier, now on display in the local ceramics museum. In the Middle Ages, around 1100,

ceramics production was revived by the Arabs, who introduced glazing and polychrome. The name 'Caltagirone' is itself presumably derived from the words *Qal'at* (hill) and *giarrone* or *inzirone* (pot) (pots' hill). Later, Caltagirone was influenced by the styles of the various foreign rulers who dominated the island in succession (Suevians, Catalans and Aragonese from the thirteenth to the fifteenth centuries), and also by the contacts with traders from Liguria and Veneto in the sixteenth and seventeenth centuries. The motifs used are a blend of Sicilian–Arab, Arab–Norman, Spanish, and Renaissance decorative elements. In the eighteenth century, ceramics production is influenced by Neapolitan products, especially those coming from Vietri (Ragona 1991).

Three natural elements are necessary in ceramics production: earth, water and fire. In the history of Caltagirone we find edicts authorizing the digging of clay pits and the cutting of timber wood. The municipal authorities are believed to have allowed the craftsmen to make free use of the vast Santo Pietro forest. In 1790 the town Senate did not comply with a Viceroy's decree ordering the Santo Pietro forest to be divided into lots. The Senate declared that: 'An interesting product would disappear, and this would bring to ruin a respectable group of craftsmen dealing with clay, who are, at present, a useful resource not only for this town, but also for the whole Kingdom' (Ragona 1991 p. 111, our translation). Thus, water, fire and earth were made available to a community that transformed them into valuable and culture-based goods.

Other institutional interventions also affected – both positively and negatively – Caltagirone's ceramics industry. In 1432 Alphonsus of Aragon, in return for services to the Crown, granted the privilege of customs duty exemption for all commodities bought or sold by Caltagirone inhabitants in every town or Crown property throughout Sicily. This privilege contributed greatly to the development and the diffusion of Caltagirone's ceramics on the island. In contrast, in 1652, a peremptory order prohibited all extraordinary expenditure and reduced ordinary expenditure for the town; not surprisingly, this was a period of decline and creative stagnation for ceramics production (Ragona 1991).

The social and economic elements necessary for the historical start-up of a cultural district can be divided, broadly speaking, into three categories: emergence of technical know-how and relative technology; professional skills; and demand. The technology used for the production of artistic ceramics, whether for a single item or a limited series, is elementary and constant: a potter's wheel and a kiln. The skilfulness constitutes the cultural element and the identity of the product, finding its aesthetic expression in the objects' shapes and decorations. Such skills, as we shall explain below, can be transmitted only through tacit knowledge. In the ceramics workshops many family members operate together: father and son, brothers, father-in-law, brother-in-law. It is on record

that in Caltagirone a number of families have worked for centuries, handing down the artistic skills from one generation to the next. Moreover, in the seventeenth century Caltagirone craftsmen producing majolica established guilds and confraternities. All the members worked together, for free, to decorate and lay the floor tiles of the important public and sacred buildings. Thus, they had the opportunity to emulate one other and to exchange technical know-how, ideas and decorative motifs. The combination of familiar tradition and shared work experience disseminates information, creating something similar to Marshall's 'industrial atmosphere' (Marshall 1920), according to the pattern of the district. The third idiosyncratic element is the demand for ceramics, largely created by the aristocratic families and the higher clergy.

At first, Caltagirone's ceramics were related to the production of honey, for which it supplied the necessary pots. The demand on the part of the aristocrats is documented by the surviving testaments and marriage contracts: throughout the seventeenth century Caltagirone's ceramics were an ever-present item in a dowry, both locally and in several towns in the hinterland. Until the end of the fifteenth century, Caltagirone was one of the most important towns of the Noto valley (Val di Noto). Its wealth consisted above all of its large feudal estate, its acknowledged Crown property, and also because noble and powerful Catalan families chose to live there, bringing with them their retinue, including ceramists–craftsmen (Ragona 1991).

The aristocrats, the dominant class in terms of political power and economic wealth, laid down the law with reference to fashion and etiquette, that is, the *savoir vivre*: from clothes to culture (music, painting, poetry), to tableware and interior decoration and furnishings. Their social status was expressed through consumption of the arts and lifestyles characterized by luxury, magnificence and excellence (Elias [1969] 1982). The aristocratic system is highly emulative, and often ruled by the economics of gift giving. This system is a means of passing on new ideas emanating from the royal courts. The desire for luxury goods made the aristocrats a major source of local demand for ceramics.

The ecclesiastical hierarchies controlled the souls of the people, but they were also feudal landowners and amassed considerable wealth as a result of taxes and revenues. They were another source of demand for sacred and secular embellishments. The demand on the part of the higher clergy consisted above all in ceramic floors, and decorative coatings for bell towers and church façades. It can be argued that the production of majolica tiles by Caltagirone's ceramists in the seventeenth century was stimulated by this ecclesiastical demand. Unfortunately, only documentary evidence remains, since the 1693 earthquake destroyed most of the buildings in the Noto valley.

These two social classes are the fundamental warp and weft for the emergence and development of a localized culture. It would be unfair to consider

them as merely passive consumers. Their cultural influence was deep and they brought in external influences, thanks to exchanges and studies carried out in other places and in other cultural environments.

In short, we believe that the existence of natural resources and key social classes (nobles, the clergy, traders and knights) in a given area is the basis for the emergence of an idiosyncratic culture, peculiar to a community, which can be handed down through the generations. Over the centuries, following economic and institutional development, the site, the local community and the whole society accumulate the cultural capital produced by the ruling classes, and consolidated through the generations.

Culture as an idiosyncratic good

Culture-based goods are idiosyncratic not only because of the material and social origin of their production, but also because of their material, intellectual, technological and commercial characteristics (Santagata 2002a). Since they are time and space specific, they possess one quality which is exclusively their own and is not subjected to the logic of competition. Historically, creativity in culture-based goods is *per se* the specific and original product of a given generation. In painting, industrial design, film-making, fashion and also in the shaping and decoration of ceramics there are generational and time-specific waves (Santagata 2002b). Generational waves are the result of the idiosyncratic character of the production of culture-based goods: the space–time continuum is crucial in creating the image and reputation of a generation.

Cultural districts are characterized by the production of idiosyncratic goods based on creativity and intellectual property. The inspiration is drawn from the cultural connection with the native local community. This strong connection with the social environment and its historical development is the source of a discriminating competitive advantage, because of the accumulation of cultural capital. Cultural or technological information is disseminated freely: it is conveyed through tacit knowledge or a tacit communication system (Polanyi 1953; Polanyi and Prosch 1975), since there is a gap between technology, art, culture and the essential facts of the real experience. Cultural goods are idiosyncratic because tacit knowledge is necessary in order to create, produce and distribute them; and also because personal knowledge is based on the individual previous idiosyncratic experience.

Intangible culture, cultural capital and the theory of institutional cultural districts

Culture-based goods, that is, goods based on the intangible cultural capital of a community or of a geographical area are naturally connected to the concept of intangible culture (in Italian: *cultura materiale*; in French: *culture*

immaterielle). The inversion of the adjective '*materiale*' in Italian and intangible/*immaterielle* in English and French is evidence of the strong connection existing in Italy between cultural production and the needs of everyday life, including leisure. The intangible culture becomes the anthropological expression of the traditional know-how that can be passed on and that is necessary to productive activities (both industrial and handicraft production). There is a common trait in intangible culture, whether we are referring to techniques, costumes or rites for agriculture, wine production, handicraft production or cultural services: it is always an intangible idiosyncratic asset.

When the intangible culture or intangible cultural capital becomes a factor of production, the character of the district becomes a key element, since the intangible localized culture is a public good, accessible to all, without exclusion costs, barriers to entry, or costs for gathering information. Information and ideas are free and easily accessible. Many small manufacturers who try to capitalize on this realize that cooperation is an advantage. This applies in particular to institutional cooperation and to all those arrangements that increase social capital.

Rights, trademarks, consortiums, cooperative banks, chambers of commerce, collective exhibition centres, promotion of local handicraft products and tourist promotion become the cement of the district industry, and an incentive to accumulate collective reputation. Thus, a long spontaneous development continues in an institutional phase characterized by the acquisition of property rights and of a sense of belonging to a community.

In fact one significant difference between the theory of industrial districts (Bagnasco 1977; Trigilia 1986; Becattini 1989) and the theory of cultural districts (Santagata 2002a, 2004) consists not only in the fundamental and pervasive presence of the local culture characterizing the cultural district, but especially in their greater proclivity towards the use of collective property rights to initiate a new phase of economic performance for the district.

Industrial districts, on the contrary, are usually characterized by a long incubation and development, whose success is deeply rooted in the local social and economic setting and in its history. In this case there is no date of birth for the district, and the long spontaneous incubation is often painful in terms of sacrifices and efforts made by the local population.

The different character of the cultural districts, consistent with the production of culture-based goods, is based on the assignment of collective property rights. In this case, despite the negative aspects, which we shall discuss below, it is possible to establish a system of positive incentives that will encourage local producers, protected from insidious unfair competition and imitation, to invest in equipment, in marketing and in collective reputation.

Institutions, trademarks and reputation in the ceramics district of Caltagirone

The character of the institutional cultural district, based on the introduction of intellectual property rights for locally processed goods (Santagata 2002a, 2004), derives from the extensive literature on trademarks and industrial, individual and collective distinctive marks (for a review, see OECD 2000).

From an economic point of view, a trademark is a distinctive mark used to reduce the information asymmetry that often characterizes economic agents during transactions (Landes and Posner 1987). As such, the trademark has a positive function for both consumers and producers; producers make use of a trademark only if it can enhance their reputation (Shapiro 1983).

Caltagirone ceramics stand out in the ceramics market thanks to their quality, which is a synonym of their capacity to evoke, in their shapes and decorative motifs, a localized cultural style. For this reason, they can be considered as a distinctive category of credence goods,[1] which we call 'hidden quality' goods. Hidden quality goods may be defined as those goods which consumers can assess *ex ante* – though a level of uncertainty remains – because of the high cost of gathering information. Such cost declines with time, thanks to the experience acquired through repeated purchases and the acquisition of knowledge. The cost of gathering information may also decrease thanks to distinctive marks which, similar to credence goods, are the evidence of their special intangible quality.[2]

The necessary steps to obtain a distinctive mark not only by private actors (the ceramists, in this specific case), but also by the institutional bodies (local authorities, museums, vocational schools) that are in charge of the protection of localized cultural wealth.

We can distinguish two kinds of distinctive mark, depending on whether they refer to a single producer or to a community of producers. The former consists in an individual distinctive mark, the latter a collective one. According to the logic of cultural districts, both kinds can be present at the same time, but only the collective or community one proves essential for the start-up of a local business system.

In the following subsections, after outlining the economic position of the Caltagirone district, we shall discuss the institutional issues, with specific reference to the situation described in the survey of the business system in Caltagirone.[3]

The economic issues

The last census of industry and services took place in 1996. Of the 740 people employed in the industrial sector in Caltagirone, 266 were working in the ceramics sector, which implies a specialization index of 36 per cent. In our survey, we found that in 2001, 327 people were working in the same

sector, indicating that this sector has taken on redundant workers from other sectors and young people looking for first employment. More than 40 per cent (135/327) of those employed in this sector belong to the family of the firm's owner, and the employees' age is quite low: more than 50 per cent are between 18 and 25.

From 1991 to 2001 the number of firms grew rapidly (+15). It is important to underline this characteristic feature of industrial districts: 52 per cent of the firms started their activity in the last decade and only three of them have existed for more than 40 years. Moreover, 66 per cent of the owners are younger than 45 years of age, and 36 per cent are younger than 35.

Ceramics production can be subdivided into mass handicraft production (using moulds), quality handicraft production (handmade, decorative work), and artistic production, where the ceramist's work is mainly creative. Generally, ceramists are engaged in two of the three kinds of production: 26 of them state that they are engaged in both quality and artistic production, 22 that they are engaged in both mass and quality production, and four are in both mass and artistic production.

The most traditional local products are ornamental pots and dishes, and tiles. The variations in price per unit are remarkable. In 2001 the price of ornamental pots varied from €14.5 to €450, ornamental dishes from €14.5 to €500 and tiles from €3.5 to €13.5.

For their finished products, firms turn to external units supplying clay and 'biscuit'. The markets supplying raw material and semi-processed goods can be local, regional and/or national; only 19 per cent of the firms turn to local markets exclusively. The data referring to sales revenues for the last three years are not very reliable, but according to the interviews, the trend was not negative and, despite some caution, sales have even increased a little.

Craftsmen–entrepreneurs rarely work for other ceramists (73/93 denied any form of cooperation). Each firm has its own laboratory (83/93), consisting of one room (34 per cent of the cases) or more (22 per cent). The laboratory is often rented and in 48 per cent of the cases it is situated in the historical centre. The mean size of the laboratories is small: 172 square metres.

The institutional background

Some of the main problems perceived by the ceramists of Caltagirone cultural district relate to institutional issues. Most ceramists would like to protect Caltagirone tradition from counterfeiting, unfair competition, and low quality due to the entry of inexpert ceramists. Therefore we might expect that they would be in favour of introducing property rights and local collective trademarks. However, there is strong resistance, resulting from excessive individualism and little understanding of the role and meaning of a collective trademark.

When asked: 'Are you interested in creating a collective trademark?', 36 ceramists answered no, 44 yes, and 13 did not answer. Nonetheless, 66 per cent (67) of the ceramists agreed on establishing a collective consortium for the protection of the trademark 'Caltagirone's artistic ceramics'. This means that more of them (two-thirds) would accept a consortium aimed at promoting Caltagirone's ceramics by means of a still undefined distinctive mark, than would agree to taking the necessary steps to create a trademark. There seems to be a contradiction here: they agree on the idea of protection, yet there is scepticism about the necessary institutional instrument required to obtain it; the trademark is seen as a useful tool for the consumer to recognize the quality of Caltagirone's products, yet it is considered more or less worthless: 'because it would benefit low-quality producers', 'we just don't need it', or 'because we lack cooperation'.

The ceramists are poorly informed about trademarks: 58 out of 93 state that a collective trademark does not exist; 17 consider it necessary. In fact, in the survey it is evident that ceramists find it difficult to cooperate and that their relationships are characterized by diffidence. They acknowledge that Caltagirone's ceramics should be protected against counterfeiting from outside the district (78/93), and they are aware that such counterfeiting exists (80/93), yet they fear even more the competition from ceramists in their own town (51/93). They accuse one another of an unspecified 'unfair competition', and fear that excess supply and low quality may bring down prices. There are also those who consider ordinary market rules as unfair competition: for example, competing to attract the best-skilled workers, the decorators, by offering higher salaries.

Individual and collective property rights

Currently, Caltagirone's ceramists sign their products, giving them a kind of unregistered individual trademark. However, this kind of distinctive mark is of little account in the market, since the commercial success of Caltagirone's ceramics is due more to the reputation of the place rather than to any single ceramist's reputation. On the basis of the interviews carried out, we learned that producers consider themselves as craftsmen–artists more than as craftsmen–small entrepreneurs; moreover, they consider their products to be original, although they acknowledge that they work within a codified tacit tradition, to which they are indebted. A signature is a kind of moral right, an individual distinctive mark strengthening the producer's artistic dimension, and enabling him/her to freely express his/her own creativity and to try to acquire a personal style and individuality within a consolidated tradition. When asked to choose between the possible creation of a normal trademark or that of a mark which represents only their own signature to protect their products, the majority of ceramists (79 out a total of 93) answered that they would rather

use their own signature and only four of them deemed it useful to use a double trademark: a logo and the signature. At any rate, the interviewees are dubious about the effectiveness of their signature as a means of public protection. Moreover, about 40 per cent of them are not even interested in obtaining more protection by establishing a register of ceramists' signatures, and 27 per cent expressed no opinion on the matter (Cuccia et al. 2000, question 75 of the questionnaire).

The trademark of Italian artistic ceramics and its relevance for the Caltagirone cultural district

At a national level, a number of laws and decrees have been passed aimed at protecting traditional artistic and quality ceramics. However, they have rarely been applied.

Law 188/90 on the 'Protection of traditional and artistic ceramics and of quality Italian ceramics' and the subsequent enforcement decrees (D.M.15/07/1996 n.506 and D.M.26/06/1997) aim at protecting the technical and product characteristics (decorative styles, shapes, quality) of traditional Italian artistic ceramics (art. 1, L.188/90). Such a goal is pursued by establishing special bodies, such as the *Consiglio nazionale ceramico* (National Ceramics Board) and the *Comitati di disciplinare* (regulations committees), and by giving this same task to regional and local authorities (each of them within its field of competence), as well as to producers' voluntary consortiums, when they exist. The chosen instrument is a protected designation of origin that can be used by those who are enrolled on the register of ceramics producers and who abide by the production regulations. The regulations define the basic characteristic of artistic ceramics with reference to models, shapes, styles and decorative motifs that are considered typical, to the processing procedures, and to the raw material used and its origin. The process is decentralized in order to prevent the trademark from becoming an instrument by means of which entry into the market can be arbitrarily held back. Nevertheless, adoption proceeds at a slow pace. Twenty-seven Italian municipalities have introduced regulations to this purpose. Caltagirone's regulations – which were introduced by the National Ceramics Board in 1996 – list seven decorative styles[4] to be protected; however, to date no producer has asked to use the trademark and the survey shows that few of them (22 out of 93) even know that it exists.

The reasons for this 'failure' are clearly economic. The cost of adopting the trademark is very high because of the strict criteria applied to select the protected products. Workshops have to conform to technical and safety standards that were conceived for larger firms. To comply with the regulations the firms should employ 'regular' workers, while at present this is mostly 'underground' labour.

To obtain the right to use the trademark, products must be decorated according to the styles that have explicitly been considered worthy of protection and valuable from an artistic and historical perspective. Thus, the ceramist's creativity is limited, and only the 'copies' of antique ceramics (those sold in antique shops) are considered to be traditional artistic ceramics. According to the law, the regulations, in order to overcome this problem, should define the assessment criteria for 'those innovative forms which represent the natural development and modernization of traditional models, techniques and styles' (art. 8, para. 2, L.188/90). Such a provision, however, is a source of problems, rather than of solutions; for if taken literally, it extends the protection according to virtually any subjective criteria.

At this point, it is necessary to try to understand why – in spite of the fact that ceramists are demanding controls aimed at protecting and promoting the production of quality ceramics – the initiatives undertaken so far have not been carried out effectively.

A model for the management of ceramists' property rights

Ceramists' behaviour with reference to the creation of a collective trademark may be explained through a game theory approach.

Let us assume that there are two types of ceramist in the ceramics market:

1. Ceramists–artists, oriented towards the production of high-quality products (art oriented). These ceramists invest in specific capital, that is, researching into and perfecting an individual and completely personal aesthetic language. They see their products as the outcome of artistic and functional research. To them high quality is essential.
2. Ceramists oriented towards the production of 'standardized' low-quality ceramics, who are interested exclusively in maximizing their profits (market oriented). Their strategy often consists in selling large quantities of low-price products. They are oriented towards industrial mass production rather than towards unique pieces. As a result, their quality is lower than the minimum level necessary for using the collective trademark.

Let us assume that an agent who can choose between the two strategies represents each category: 'adopting' or 'not adopting' the collective quality trademark. The outcome of the interaction in a static game with simultaneous strategies, complete information and non-cooperative behaviour – that is, the most appropriate model in this research, in our opinion – is presented in the following pay-off matrix (Table 21.1), where C_A stands for the representative of art-oriented or 'high-quality' ceramists, and C_M stands for the representative of market-oriented or 'low-quality' ceramists, who would not respect the minimum standards set by the collective distinctive mark.

Table 21.1 Pay-off matrix of the game

		C_A	
		Adopting	*Not adopting*
C_M	*Adopting*	$\pi + \varepsilon$ $\pi - \varepsilon$	π $-\pi$
	Not adopting	ε π	0 0

If player C_A decides to adopt the trademark, his/her pay-offs are the following: π if player C_M does not adopt it, or $(\pi - \varepsilon)$ if C_M adopts it, thus creating a negative externality, as the mean quality level which the consumers associate with that trademark falls.

On the contrary, if player C_A decides not to adopt the trademark, his/her pay-offs are:

- 0 if C_M does not adopt it as well (this may be interpreted as the normalized profit level, which is equal to a situation where there is no trademark),
- $-\pi$, that is, a negative profit, if C_M adopts it, thus benefiting from the advantages a trademark may offer on the market.

As regards C_M, if he/she decides to adopt the trademark, his/her pay-offs are the following:

- $\pi + \varepsilon$ if also C_A adopts it (in this case C_M, at no cost, benefits from a positive externality, ε, that is, the higher mean quality associated with the trademark thanks to the presence of C_A), and
- π if C_A does not adopt it.

If C_M decides not to adopt the trademark, his/her pay-offs are:

- 0 if C_A does not adopt it, and
- ε if C_A adopts it (although he/she did not adopt it, C_M can benefit from the externalities arising from his/her being a ceramist in the same area with a solid tradition of ceramics).

The matrix of the game clearly shows that the representative player C_M has a dominant strategy: adopting the trademark. Whatever C_A's behaviour, C_M

will find it convenient to adopt the trademark, since $\pi + \varepsilon$ and π are greater than ε and 0, respectively. This raises a well-founded and serious dilemma in ceramists–artists, who produce high-quality goods.

The representative player C_A may have a dominant strategy if and only if ε is low ($\pi - \varepsilon$, π are always better than $-\pi$ and 0, respectively). More precisely, by low ε we mean $\pi - \varepsilon > -\pi$, that is $\varepsilon < 2\pi$. If ε is high, C_A has no dominant strategy. In that case the iterated dominance criterion is usually adopted: if C_M plays his/her dominant strategy adopting the trademark, C_A will play, respectively, adopting if ε is low (Nash equilibrium will then be adopting–adopting) and not-adopting if ε is high (Nash equilibrium will then be adopting–not adopting).

We have a small negative externality (low ε), for example, if both types of ceramists (market and art oriented) share a trademark which guarantees only the geographical origin of the product (appellation of origin) without linking the geographical origin with specific quality standards. Each ceramist – through his/her signature or a personal trademark in addition to the collective one – may then differentiate and characterize his/her own products from a qualitative point of view. The collective mark guarantees only the minimum quality.[5]

On the contrary, we have a big negative externality (high ε) if the collective mark guarantees not only the origin, but also the product's mean quality. In that case, if low-quality ceramists adopt the trademark, it will gradually become a signal of decreasing mean quality.[6]

If we analyse the Pareto efficiency of the Nash equilibria, we note that: (a) if $\varepsilon < 2\pi$, Nash-equilibrium adopting–adopting is also Pareto efficient; and (b) if $\varepsilon > 2\pi$, Nash-equilibrium adopting–not adopting is not Pareto efficient, as compared to the not adopting–adopting outcome.

By repeating the game an infinite number of times and by applying the Folk Theorem (Friedman 1971), we have: (a) if $\varepsilon < 2\pi$, the efficient equilibrium adopting–adopting is repeated an infinite number of times; and (b) if $\varepsilon > 2\pi$, repeating the Pareto-efficient outcome not adopting–adopting, may become a Nash equilibrium. This depends on the discount rate, that is, on the level of C_M's impatience.

The condition for the infinite repetition of the Pareto-efficient solution to become a Nash equilibrium of the repeated game is the following:

$$\pi < |\varepsilon - \pi| \cdot \sum_{t=1}^{\infty} \left(\frac{1}{1+\rho_M} \right)^t$$

where π is the short-term gain of profit due to C_M's choice of adopting the trademark rather than not adopting it (note that $\pi = \pi + \varepsilon - \varepsilon$); $|\varepsilon - \pi|$ is the loss in each future period, if C_M has chosen to adopt the trademark in the

present period and therefore C_A replies by not adopting it in each of the future periods, ρ_M is the discount rate applied by C_M to future profits.[7]

The condition may also be written as $\rho_M < |\varepsilon - \pi|/\pi$. Therefore, the representative player C_M prefers to play 'not adopting', which brings about a Pareto-efficient equilibrium, if and only if his/her discount rate ρ_M is lower than the threshold value ($|\varepsilon - \pi/\pi$). In this case, the market-oriented ceramist C_M proves to be patient and attributes more importance to the future profits arising from not adopting the trademark than to the present profits arising from adopting it.

If market-oriented ceramists make their choice in a long-term perspective and apply to their future profits a rather low discount rate (lower than the above-mentioned threshold), they might also decide not to adopt the trademark because, once the trademark has become established on the market and is considered indicative of high mean quality – thanks to the ceramist–artist – that may also produce positive externalities for those who work in the same area but on lower-quality market segments (for example, the typical mass-produced souvenirs for tourists who are nonetheless attracted by the trademark of artistic quality). Thus, choosing not to adopt the trademark today might mean higher profits in the future.

If, instead, market-oriented ceramists make their choice in a short-term perspective, they will choose to adopt the trademark and maximize their current profits, and they will ignore the fact that the lower mean quality – due to their decision to adopt it – will lower everybody's profits in the future.

The above discussion makes it clear why ceramists are sceptical about trademarks. The creation of collective distinctive marks leads to economic efficient allocations only under specific parameter restrictions. Therefore, the creation of collective distinctive marks should be structured in the following way: if the externality is higher than a threshold value, a separating equilibrium should be generated, so as to make it possible to distinguish among producers. Consumers would then receive the correct information. In the long term, by means of the separating equilibria, ceramists–artists and ceramists–entrepreneurs may coexist on the market.

Conclusions

Introducing intellectual property rights to protect goods that are both idiosyncratic and universal, such as culture-based goods, seems to be one of the most effective tools to start up a cultural district and foster localized industry.

Caltagirone has an ancient and valuable tradition of high-quality artistic ceramics, and by introducing intellectual property rights, namely the protected designation of origin right, one might start up a new phase of development at district level. The experience in this cultural district shows us

that designing such a tool correctly is not sufficient: there are other essential aspects which may turn the initiative into a success or a failure, such as disseminating information and creating capital stock.

Every ceramist is aware that Caltagirone's ceramics must be promoted *collectively* in order to achieve a national and international reputation. However, it would seem that they are not prepared to renounce their individuality by using a trademark which might evoke, even to a small extent, a sort of industrial production.

Is such reluctance due to a diffuse lack of cooperation among the entrepreneurs of this area or can it be traced to some peculiarities of this economic sector? In this sector we can distinguish two main types of ceramist, those who consider themselves as artists, and those who are market oriented. The game theory approach demonstrated that the market-oriented ceramists, if they work in a short-term perspective, might prefer to use a collective trademark for their low-quality products and thus induce the ceramists–artists not to use it, as it would no longer be a sign of high quality on the market and a form of protection for the local products. The distrust and scepticism of Caltagirone's ceramists towards collective tools aimed at protecting intellectual property may find a theoretical justification.

Nevertheless, as local tradition is expressed through non-material inputs (that is, creativity and artistic decoration), it is highly vulnerable, as these factors are mobile and can combine with other necessary inputs for that production in other areas, thus dispersing the value arising from the traditional origin of goods. It is a matter of urgency to protect this space-specific wealth. But the creation of a trademark should represent only the beginning of a process aimed at making public and private institutions aware that they operate in an area that is highly specialized in traditional production and could become a 'cultural district'.

Notes

* This research was carried out thanks to contributions by CNR, *Progetto finalizzato beni culturali*, n.01.00631.PF36, and by MIUR, *Programma operativo del piano beni culturali*, cluster 29, project n. 3–T3. T. Cuccia, G. Cucuzza, W. Santagata and G. Signorello participated in the research and in the survey. The questionnaire used is available on request. The authors thank the research team, and assume full responsibility for this chapter's content. English translation by Elena Pasquini.

1. On the basis of a consolidated taxonomy, which is widely used in industrial economics (Nelson 1970; Darby and Karni 1973), goods (or one good that can be used for different purposes) can be divided, according to the degree of possible assessment of their quality on the part of the consumer when he/she buys it, into 'search goods', whose quality can be ascertained *ex ante*, before purchase; 'experience goods', whose quality can be ascertained only *ex post*, after purchase; and 'credence goods', whose quality cannot be ascertained, even after purchase.
2. We refer to certified organic products and to 'environmentally friendly' goods.
3. Of a total of 120 firms registered in the year 2000 in the *Albo Imprese Artigiane della Camera di Commercio di Catania* (Register of craftsmen–entrepreneurs of the Catania

Chamber of Commerce), 112 agreed to answer the questionnaire and were interviewed at their workshop; 93 questionnaires were considered reliable for data processing.
4. The seven protected Caltagirone artistic styles are: proto-majolica, majolica typical of Chiaramonte, and the decorative styles of the fifteenth, sixteenth, seventeenth, eighteenth and nineteenth centuries.
5. For example, a collective mark such as a geographical indication (in Italy, *Indicazione Geografica Territoriale*), which is granted to those products whose quality is mostly or completely due to the geographic area indicated, but are only partly processed in that area, and only a part of their contents must necessarily come from that area.
6. For example, a collective mark like the protected designation of origin (in Italy, *Denominazione di Origine Controllata*), which is granted to those products whose quality is essentially or exclusively due to the geographical environment of origin and whose material and non-material content come entirely from that geographic area.
7. $\pi - \varepsilon$ is negative because $\varepsilon > 2\pi$, therefore lost future earnings are $\varepsilon - \pi$.

References

Bagnasco, Arnaldo (1977), *Tre Italie. La Problematica Territoriale dello Sviluppo Economico Italiano*, Bologna: Il Mulino.

Becattini, Giacomo (ed.) (1989), *Modelli locali di sviluppo*, Bologna: Il Mulino.

Benghozi, P. and W. Santagata (2001), 'Market piracy in the design-based industry: economics and policy regulation', *Économie Appliquée*, **3**, 121–48.

Cuccia, T., G. Cucuzza, W. Santagata and G. Signorello (2001), 'Indagine sul distretto culturale industriale della ceramica di Caltagirone. Il Questionario', unpublished.

Darby, M. and E. Karni (1973), 'Free competition and the optimal amount of fraud', *Journal of Law and Economics*, **16**, 67–88.

Elias, Norbert (1969), *Über den Prozess der Zivilisation*, Italian translation: *Il Processo di Civilizzazione* (1982), Bologna: Il Mulino.

Friedman, J.W. (1971), 'A non-cooperative equilibrium for supergames', *Review of Economic Studies*, **38**, 1–12.

Landes, W. and R. Posner (1987), 'Trademark law. An economic perspective', *Journal of Law and Economics*, **30** (2), 265–309.

Marshall, Alfred (1920), *Principles of Economics*, 8th edn, London*:* Macmillan.

Nelson, P. (1970), 'Information and consumer behaviour', *Journal of Political Economy*, **81**, 729–54.

Organization for Economic Cooperation and Development (OECD) (2000), 'Appellations of origin and geographical indications in OECD member countries: economic and legal implications', unpublished.

Polanyi, Michael (1953), *Personal Knowledge*, Chicago: University of Chicago Press.

Polanyi, Michael and Harry Prosch (1975), *Meanings*, Chicago: University of Chicago Press.

Ragona, Antonino (1991), *Terracotta – La Cultura Ceramica a Caltagirone* [Earthware – The Ceramics Culture of Caltagirone], Catania: Domenico Sanfilippo editore.

Santagata, W. (2002a), 'Cultural districts, property rights and sustainable economic growth', *International Journal of Urban and Regional Research*, **26**, 9–23.

Santagata, W. (2002b), 'Some effect of creativity on fashion market behavior', paper presented at the International Centre for Economic Research (ACEI) International Conference, 2002, Rotterdam, Working Paper EBLA CENTER, www.de.unito.it.

Santagata, Walter (2004), 'Cultural districts and economic development', in Victor Ginsburgh and David Throsby (eds), *Handbook on the Economics of Art and Culture*, Handbooks in Economics series, edited by Kenneth Arrow and Michael D. Intriligator, Amsterdam: Elsevier Science, North-Holland, forthcoming.

Scott, Allen J. (2000), *The Cultural Economy of Cities*, London: Sage.

Shapiro, C. (1983), 'Premiums for high quality products as rents to reputation', *Quarterly Journal of Economics*, **98**, 659–80.

Trigilia, Carlo (1986), *Grandi Partiti e Piccole Imprese* [Big Parties and Small Firms], Bologna: Il Mulino.

22 Property rights in higher education

Ryan C. Amacher and Roger E. Meiners

Introduction

Property rights are generally understood to be claims to resources that are formally or informally recognized to be under the control of a person or group of people. As most of the chapters in this volume discuss, we are generally concerned with resources that are privately held. When we move away from the world of individual decision makers and profit-driven organizations that allocate property, we are in the world of public sector control of resources or control by non-profit organizations. While the same self-interested individuals are at work in the public and non-profit sectors as are at work in proprietary organizations, the lack of profit incentives causes different results to evolve.

One large market dominated by public and non-profit organizations is higher education. Higher education would exist in the absence of public intervention, as historically it was, at least in the United States, dominated by churches (Meiners 1995). For-profit colleges command a small fraction of the market. It is hard to compete with the public purse and subsidies provided by donors to non-profit organizations. Our focus is not on why higher education is dominated by government and non-profit organizations but what we observe the incentives to be of participants in these organizations. Differences are observed between government and non-profit colleges. Self-interest is always at work but is constrained and directed by the rules of the game. The implicit and explicit property rights within colleges and universities explain much of what we observe.

Critics of higher education tend to focus on actions that reflect people responding to the incentives faced within universities. Results that many see as objectionable, whether it is waste or lack of response, often result from institutional design. To change the outcomes there must be changes in incentives.

A more fruitful approach to understanding higher education is based on property rights. To talk of students as customers who demand higher education that is supplied by universities is nebulous. A university (or college) provides students a set of rights to resources subject to conditions and responsibilities. Markets should be viewed as an exchange of rights rather than as an exchange of commodities (Demsetz 1967). While true as a general proposition, it is especially salient as we examine the exchanges that occur in

higher education. We begin by considering the governance structure of universities and then move to a discussion of the incentives of administrators, faculty and students (for an extended discussion of some points, see Amacher and Meiners 2003).

University governance

Imagine a company that has no bottom line; no profit measure to compare performance from year to year or against competitors. Unlike a for-profit firm that has a clear ownership structure and line of authority, these organizations are political or charitable at base and have vague or shifting lines of authority. The head of the organization, the president, is often selected by a committee dominated by employees. If the employees become unhappy with the president, they can demand that he/she be fired – but employees are not fired except in rare cases. The president reports to a board, but it is often passive. Its members do not have expertise in the organization's area of production and have no personal profit motive or profit signals to guide their actions. In a nutshell, that's the world of university governance. The problem is especially pervasive at state universities. The same governance structure exists at private universities but the problems need not be as pronounced because the boards of trustees are more likely to have clear legal authority to have control at private colleges and have ultimate budget responsibility.

Universities, whether state agencies or private non-profit organizations, do not have the kind of financial measures that organizations in the private sector rely upon to drive performance decisions. There are performance measures, but colleges are shy to adopt them, and often rely on little more than cash balances as measures of success. Measuring performance means increased responsibility; few people volunteer for increased responsibility unless there are rewards to go with it, so attempts to measure effectiveness by market standards are generally opposed. Employees at private companies would be likely to act with as little accountability as university faculty and administrators do, but are restrained because the market demands performance and, sooner or later, disciplines unproductive practices.

The role of trustees

Trustees of a college are like the members of a corporate board of directors. Legally, they are the principals of the organization, responsible for the successful completion of its mission. For corporations, the mission is the long-run profit maximization of the company to enhance stockholder value. For colleges, the ostensible mission is to maximize the quality of their educational output over time. Companies have financial statements and other data that can be evaluated by board members and by outsiders. When a board of directors does a poor job of hiring and monitoring top managers to enhance

firm value, the stock price is lower than the company is believed to be worth if in the hands of another set of managers. Even if the firm is not publicly traded, outsiders with expertise in the area can make expert evaluations about the relative effectiveness of current use of the assets controlled by the organization. If the current directors and managers do not correct the situation, outsiders can mount a takeover effort to gain control of the board, replace top management, restructure the company, and enhance company performance.

The history of takeovers indicates that most such decisions are justified; the firm or group that engineers the takeover generally improves the performance of the target. If a mistake is made, those who engineered the takeover suffer a financial penalty. In contrast, colleges have no financial market that allows outsiders to evaluate performance or allows a takeover to be mounted to replace a poor-quality board with a higher-quality board that demands better results from top management. Indeed, sloppy management of public colleges usually produces calls for more taxpayer money to solve the 'underfunding' problem. University trustees are usually volunteers who think it an honor to sit on the board of their Alma Mater and know that they are expected to be donors to the institution, not draw a salary. Most trustees have no expertise in higher education and are passive, allowing top management to be in control unless things seem to be getting grossly out of hand – which may be for myriad reasons.

Trustees rarely focus on the kind of objective measures available to guide corporate board members. Members of boards of for-profit organizations have a duty to analyse financial information that provides measures of managerial effectiveness and many board members have a personal financial stake in the future profitability of the firm, giving a further incentive to work for its success. In contract, colleges have financial numbers that provide less guidance for evaluators about the relative success of the organization. Unless top college administrators mismanage budgets, where board members have expertise and there are numbers to study, the most likely feedback that trustees can take into account about administrators is complaints from faculty or negative press about a real or alleged problem. To avoid such trouble, administrators have a strong incentive to keep things calm. Administrators who make waves are likely to get faculty and lower-level administrators up in arms if they are threatened by the change. If the board is not solidly in the president's camp, and is willing to bear complaints, the president is likely to be replaced by someone who is more expert at keeping the lid on.

Presidents are, therefore, primarily in charge of putting out fires and keeping trouble at bay. President Gerhard Casper of Stanford University notes:

> Many people think of universities as hierarchical because they have a president with a fancy title … but they are not hierarchical. Power comes from the bottom

> up. The most important decisions are those concerning admissions, curriculum and faculty appointments, and these are areas where the university president has almost no power. In most circumstances, I'm the man with the pail and broom. (Honan 1994, p. 16)

Unlike private companies, where presidents are hired and often given specific orders, such as eliminate losing operations, streamline management, change service lines, and so forth, few college presidents are specifically charged with cutting losing academic departments and unloading unproductive faculty. Boards are often clueless about such matters. Presidents who undertake efficiency or academically innovative moves without solid board approval are likely to be headed for a shorter term in office than if they leave things pretty much as is.

Some college boards hire presidents and charge them with hard management missions; if they are willing to back up the president during the grief that is sure to come, the president may get something done. But many boards allow themselves to be handed a list of presidential finalists by a faculty-dominated committee. Such committees have no interest in offering candidates to be considered for president if they think the candidate would be a threat to the existing order. The list is usually composed of administrators from other institutions who have succeeded – from the faculty perspective – by controlling problems rather than dealing with hard issues by forcing change. Much more so than is the case of employees at for-profit organizations, faculty often have effective control of the choice of top management of colleges. While it is true that in some for-profit professional organizations, such as large law and accounting firms, the senior employees (partners) play a significant role in management decisions, it is also the case that these organizations reallocate resources and pay based on financial measures of performance. Passive boards means that passive presidents, or at least ones not likely to disturb the current allocation of resources, are likely to be hired. The status quo reigns.

To stay in a president's job longer than the average tenure of four years in public universities, means balancing the books, talk about commitment to excellence, and do not force substantive change. A passive board finds little reason to fire such a president. Presidents who press for change force trustees to decide about controversial matters about which they have little experience. Trustees rarely think the faculty is right, but there is no financial reward in bearing the costs of a fight in which they will be called names in the media for supposedly destroying academic quality and freedom. A common solution, because troublesome faculty members are not leaving, unlike troublesome or unproductive members of a for-profit professional organization, is to force out the president and hope that the next one will avoid such disputes.

In sum, college presidents can be expected to be less entrepreneurial or creative than presidents of for-profit organizations, and can be expected to

tolerate employees who could be replaced by better employees (faculty) for the same or lower salary. A company president has financial targets to work toward; a college president need only do no worse, in financial terms, than similar public or private non-profit colleges. At state colleges they can point at the legislature for not granting all wishes in the proposed budget. There is little incentive to try to improve resource allocation or other controversial moves that may enhance a college over time – that can mean a fight with some incumbent faculty and a president is not often rewarded carrying on a fight that is likely to be played out in the media. Much more than in the for-profit sector, judgments by college trustees about the performance of presidents is based on subjective feelings about how things are going.

Differences between private and public colleges

Both public and private colleges have boards of trustees. While most private college boards are legally the principals of the organizations, and may control the college as they think best to fulfill its educational mission, public colleges are state agencies. The boards usually have power to appoint the president, but they are representatives of the state to oversee the agency on behalf of the citizens. How much authority public college trustees have is a matter of state law; in some states they have substantial power, almost like private college trustees; in other states the boards are largely ceremonial, and real power is in a central state agency, such as a higher education commission.

State colleges must follow state and federal laws governing purchasing, spending and employment. State universities (and other state agencies) often pay more for things than if they were not subject to intricate rules. Colleges may spend more on overhead costs preparing requests for bids and processing the bids than the things being bought are worth. Invoices for trivial purchases may have to clear multiple levels of approval. Such rules are usually imposed in the wake of someone doing something wrong, but efforts to avoid occasional theft or bad decisions do not justify raising the costs of all purchases. Complexity makes purchasing less, not more competitive. States also impose rules such as those mandating that a certain percentage of business must be done with minority contractors. Even if such rules are created with the best of intent, they impose a substantial cost both in price paid and overhead staff needed to ensure compliance. They can create the appearance that college administrators are not very bright, when they are simply following rules that limit administrator flexibility to control resources in the most effective manner.

Private colleges face fewer constraints on buying practices and many other operational features. They may set their tuition and fee rates, and offer scholarships to students as they see fit; that is, they face few constraints on price discrimination. The most desirable students are offered lower tuition

fees and other benefits. State colleges charge a tuition rate set by the legislature or the commission. Scholarships, unless funded by a special pool, which many states have established for minority students, must be funded by private donations. Whether it is buying computers or attracting students, administrators of state colleges have fewer degrees of freedom than do private school administrators. That is, the administrators at public universities tend to have fewer property rights within the organization than do administrators at private universities.

Evidence of this comes from the fact that public colleges have a difficult time attracting academic stars. Nobel Prize winners tend to be concentrated at private colleges that have fewer pay constraints and bargain for whatever work terms the parties desire. State colleges usually cannot go over some politically-acceptable salary level, so administrators have less flexibility in creating employment contracts than do administrators of private colleges. Public colleges usually rely on private donations to provide salary support for academic stars who could not be attracted by public funds alone. The constraints at public universities, whether imposed by state law or implicitly understood to exist by administrators who wish to keep their jobs, mean that the top-ranked colleges in the nation, both at the graduate research level and at the undergraduate level, are dominated by private colleges. Trustees at private schools generally allow administrators to be more creative than is often the case at public colleges. Public universities, to avoid possible budget punishment by the legislature, are likely to avoid unusual programs that some private colleges may adopt to appeal to a market niche.

Resource control via internal administration

The specialization of disciplines requires administrators to rely on the faculty to provide expert judgment about curriculum and faculty in their fields. The key issues are: how much control will faculty have, how will the control be exercised, and what incentives do they have? Monitoring faculty is difficult. Simple measures such as number of students taught, number of hours spent in an office, or number of publications produced may have little relationship to quality and effectiveness. The issue in education – as in most intellectual activity – is the quality of the results. To focus on input measures, which may give the appearance of being objective or scientific, can produce perverse results.

A great challenge in academic administration is to hold individuals responsible for what they produce with the resources at hand. Faculty members enjoy a significant lack of personal accountability and there are few measures of financial productivity. Responsibility can be masked by committee decisions, which is one reason colleges are dominated by committee decisions. Administrators can skirt responsibility by appointing committees and choos-

ing among the options presented by the committee. Skillful administrators appoint committees they know are likely to produce the results desired; in any event, valuable time of faculty and administrators is consumed in committees that produce decisions that easily could have been predicted. This process helps protect administrators from criticism by faculty and supervisors, but does not mean that the college as an institution is improved or even that current resources are effectively allocated.

The domination of college decision making by committees means that a huge number of faculty hours are spent in nattering discussions – time that could have been spent on other college business, such as teaching or research. This time could be justified if one expected committees to produce better decisions than individual decision makers, who would be likely to gather information informally from relevant faculty and staff. Committees tend to represent and perpetuate the status quo. Little innovation can be expected. Tomorrow looking like today is more satisfying to most people than an uncertain future, especially one that could require people to be more productive, such as prepare new course material, with no expectation of higher compensation.

It won't cost anything!

Economists are fond of teaching that there is no such thing as a free lunch. But even economists, when they become academic administrators, will assert that there are free lunches. When a new degree program is proposed, the faculty and administrators making the proposal have incentives to assert that the program will cost little but will bring in revenues and academic recognition. Since it is rare for the department or college proposing the new program to drop an existing program in its place, or for faculty to volunteer to teach more for no additional compensation so that the program can be offered, it is silly to say that a program will cost nothing. But this game is played all the time, especially at public universities that must go through an internal and external bureaucracy for permission to do things. It is one way to play the game to try to capture more resources, which may have little to do with academic merit either in terms of quality or of what is demanded by students.

At many universities, faculty representatives from around the university can vote on a proposal by, say, the nursing college to, say, begin a master's program in pediatric nursing. To ensure continued control over existing resources that faculty and administrators think of as their property, other colleges want assurance that the university will not devote a larger share of its budget to the college proposing a new program. It does not matter if the program has a large number of demanders and is likely to be very successful; a master's program in pediatric nursing is likely to be opposed by other colleges if they think it could cost them resources. Faculty and administrators tend to think of

college budgets as a fixed pie to be sliced up rather than as a pie to be made and expanded by contributions by all those who are supposed to help bake the pie.

This unproductive control of resources within universities is further encouraged by policies in many states that require colleges to go to the state capital to talk to state commissions about new programs. Whether or not a proposed program may have merit matters little and, in most cases, is beyond the ability of the bureaucrats to judge. Commissions use measures such as 'duplication' within the state, what the cost will be, and whether or not there is political support for the proposal. To reduce the cost objection (the bureaucrats can usually only measure inputs or costs, they have little incentive or ability to measure outputs or benefits), colleges may claim that new programs will be 'free'. Administrators assert that they will shuffle existing resources and that the new program will cost nothing. If they tell the truth about program costs they are less likely to be successful in getting approval, regardless of possible future success. The incentive structure means that there is a scramble for extending control over resources, or inputs, with little relationship to outputs, although great concerns are expressed about outputs. In bureaucracies, once resources are allocated, they effectively become the property of those who control them, regardless of the relative merits of alternative resource use.

Reducing competition

The concern for program duplication has some merit, given the current method of state appropriations, but would have little merit if public colleges, like private colleges, had to compete rather than be granted protection by the state cartel manager, the state commission on higher education. No doubt restaurants, physicians, or any other service provider would like a state agency to decide when there was 'enough' of something being provided. Burger King would always argue against allowing a rival operator to open a franchise nearby and physicians would argue against allowing more doctors to open their office in towns because there are 'enough' doctors.

A state college that has an engineering school would prefer to be the only one in the state since it would have a lock on many students and would face less competition. Since the state is footing the bill, and the primary measures are inputs, there is reason to be wary of the value of new programs proposed by state colleges. On the other hand, any program the state will pay for is worth asking for from the perspective of university administrators, even if the program makes little sense academically. If the 'experts' in the capital want more engineering programs to supposedly boost economic development, proposals will come flooding in from state colleges for engineering programs. The focus of state college administrators is on establishing a claim on state

resources; few measures of the effectiveness of the use of the resources are available.

Share and share alike

A favorite topic of some faculty is 'shared governance', which means 'democratic' control of a university by faculty. As Harvard Dean Rosovsky notes, at a university 'more democracy is not necessarily better' (Rosovsky 1990, p. 265). Contrary to what many faculty think, they are not hired to determine general university policy. 'Faculty members are invited to teach and do research and to set educational policy in their sphere of knowledge' (ibid., p. 266). Faculty have little expertise in academic areas beyond their own, so have little to contribute to decisions in other areas. Furthermore, there is a conflict of interest between the personal interests of faculty to command control of resources (and even more so of students) and what may be the long-term interests of a college.

Colleges and departments within colleges that are democratically controlled will tend to represent the 'average' faculty preference. This is seen most easily in pay-raise matters. Left to a faculty vote, the majority will likely favor an across-the-board pay increase; everyone gets the same percent (or lump-sum) raise. Administrators would rather have leeway to give less productive faculty nothing (which may encourage their departure or convince them to work harder) and the most productive faculty something extra.

Colleges where a union represents the faculty, which means little administrative leeway in pay decisions, are lesser-quality institutions with faculty members who have fewer employment alternatives. No high-quality university has a unionized faculty. Highly productive people with good alternatives do not want to be trapped in an equal-share university. Even if a faculty is not unionized, empirical evidence confirms the commonsense notion that the more power the faculty has in administrative decision making in a college, the lower the quality of the college (McCormick and Meiners 1988).

When average faculty preferences dominate college decisions, resources are likely to be allocated on an 'equitable' basis, such as everyone gets their way paid to two professional meetings a year. Professor A, highly regarded in a field, may be invited to go to numerous major meetings, while his colleague Professor B, who produces nothing, volunteers to give a couple of minutes worth of comments at low-level meetings. With the same resources to support four trips, a sensible administrator would prefer to give A three or four trips and B one or none. The college's reputation is enhanced more by having Professor A appear at more meetings, Professor A is less likely to be enticed to another college that offers more support for such events, and good work is rewarded. Administrators who discriminate on the basis of productivity face the wrath of less-productive faculty. If things are done 'democratically', the

administrator had better have solid backing from higher administrators or trustees to support discrimination in the distribution of resources because those who do not get a 'fair share' may cause endless problems.

Similarly, shared resources, such as secretarial assistance, are usually distributed on an equality basis; everyone has the same right of time usage, so productive faculty do not get enough help, while the unproductive faculty waste secretarial time having trivial matters addressed. An irony of computers is that the most productive faculty have tended to move to producing more of their own work; unproductive faculty are more likely to rely on costly secretarial help for their meager output. Similarly, requests to the library for new books are treated the same whether they are from a productive professor, who actually reads and uses books, or from a slacker who had not read a book in years, but goes through the motions of being intellectually alert.

Everyone is valued equally

An example of faculty democracy that produces results that are inconsistent with the support of the advancement of knowledge is in how many colleges distribute internal research grants. Unlike the National Science Foundation, which, as inefficient as government allocation of research dollars may be, at least uses reviews by authorities in the fields in which grants are being requested, many universities have a university-wide faculty committee distribute grant money to faculty who apply. The bizarre spectacle then exists of economists and historians sitting in judgment of the relative merits of proposals from faculty in pediatric nursing, chemistry and sociology. Almost no one on the committee is competent to judge the merits of the proposals, yet the members solemnly make pronouncements about subjects about which they know less than many students. The property rights structure within universities ensures relatively inefficient use of resources over time.

When employees (faculty) determine university policy by so-called democratic decision process, the most unproductive member of the faculty has the same weight as a Nobel Prize winner. Hence, faculty members of below-average quality lobby hard for more and more things to be decided by democratic votes, and to take decisions to committees, away from administrators.

This problem is compounded because faculty members who lobby for committee assignments and for election to large committees such as the Faculty Senate, are rarely the best faculty. The best faculty have higher opportunity cost of their time for teaching, writing, and involvement in external activities. They are the most mobile in the job market. Competent faculty have no desire to spend afternoons trapped in conference rooms with colleagues who spend hours making pompous pronouncements about any issue that arises. Faculty committees tend to represent those who have good reason

to lobby to keep the faculty in control of as many resources and decisions as possible. Administrators know that these faculty members will spend countless hours fighting trivial changes in resource allocation so they get their share of any resources.

Faculty resist change at a university because it can mean more work and greater uncertainty about the future. Especially at state colleges, where the state can be relied upon to send support year after year to maintain the status quo, the safest bet is for faculty to support administrators who leave things alone and, even better, spend their time raising extra money to support expansion of existing programs. As the head of the Institute for Research on Higher Education at the University of Pennsylvania notes, 'senior professors lack any financial incentive to support [change]' and see no reason to cooperate with anyone who threatens life as it is, because nothing extra is being offered, so their attitude is 'why cooperate … I'm outta here in 10 or 11 years … so why should I bother?' (Honan quoting Zemsky, 1994, p. 18).

Administrators' lives are easier if they give in to the desire by many faculty for tomorrow to look like today. As Don Paarlberg, a leading agricultural economist, has noted, many parts of agricultural colleges make no sense given the modern structure of agriculture, yet the federal government and the states keep pumping money into academic departments with very few students (Paarlberg 1992, p. 45). Supported by such political largess, agriculture colleges even have programs in agricultural business that should be a part of regular business education, just as agricultural biology departments should be merged into regular biology departments. We know of schools that have more tenured professors in some agriculture departments than there are undergraduate majors. But what would make for a more efficient reallocation of resources in terms of what students demand is irrelevant when colleges can get funds for programs regardless of whether or not students are enrolled.

The problem of faculty incentives

Administrators who take actions that would be normal in a for-profit organization, such as dismissing the worst tenured faculty members and abolishing poorly-performing programs, send chills through the ranks of everyone who worries that they are or might become deadwood. If faculty want assurances that tenure means a life sinecure, which is *not* what tenure grants legally, they have strong incentives, usually under the guise of academic freedom, to protest any move against incompetent faculty. Secure property rights in employment are valued by most faculty, so they have little incentive to support administrators who would like to terminate the employment of the least competent as it could mean insecure employment rights for all. Unlike in a for-profit organization, where bad employees can cost everyone money and it is understood that employment rights are insecure, at a college, an unproduc-

tive professor is unlikely to lower the income of another faculty. It is always comforting to know that, regardless of how hard (or not hard) you work, you are very unlikely to be subject to dismissal. Competition is a hard world best avoided if possible.

Given the limited control administrators have over faculty, faculty evaluation is a touchy business. As at any place of employment, no one enjoys being compared to other employees (unless you always rank number one). If productivity can be measured by sales, number of calls made, number of widgets produced, the number of clients generated, and so forth, there is little room for disagreement – the highest producers get paid the most. But how does one decide rewards among a group of professors?

The problem with good teaching is to give faculty an incentive to do a good job consistently. Private colleges, in general, demand better teaching than do public colleges, but the management problem is common to both sets of institutions. Unless faculty income is tied to substantive teaching performance, there is little incentive to invest time and effort in high-quality teaching. Even if good teaching makes students happy, how can faculty capture any of the gains other than the value of popularity?

Once a faculty member earns tenure at a university, good teaching is hard to capitalize into pay. Time spent on committees that try to grab college resources may be time well spent (not for the university, but for the faculty members), and time spent producing published research can be capitalized into pay by staying active in the job market. Faculty who spend lots of time politicking have little time for research and writing; as noted before, committee work and on-campus politics appeals most to faculty who have, on average, less ability to publish their way to higher pay raises. Faculty publication is a good thing; the issue is the *balance* between teaching and publication. Colleges do not want professors to maximize publications by shirking on teaching quality.

Unproductive faculty have weak publication records, so are unlikely to be offered another academic job. They are most likely to invest their time maximizing the benefits of their current position, which means fighting for resources and resisting change. Administrators who cross them must to do battle with faculty who have nothing but time on their hands to spend lobbying against proposals for change. When a scrap occurs between administrators and some faculty, the public often just sees a professor who claims to be crusading for academic freedom or talks about serving students. The media and outsiders do not know one physics professor from another and, therefore, do not know if the least competent faculty are leading the charge against administrators. It always makes good news copy to have an outspoken professor attacking 'waste' by an administration that 'does not care' about students.

Administrators who want an easier life do not take on the worst-performing employees: they make peace instead. Faculty generally think well of

administrators who do not threaten the status quo or who devote their time to fund raising rather than trying to enhance the value of resources at hand. Hence, trustees have a hard time discerning whether administrators are liked by faculty because they are so skillful that productivity improvements can be made without upsetting people, or they are popular because they are not doing much. If trustees guided for-profit institutions, with better measures of productivity, they would not have to rely on such weak forms of input to help guide decisions.

Institutional consequences

Universities, dominated by the government and non-profit status, suffer more efficiency problems than would be expected in for-profit firms. This is nothing new. Those who think colleges have 'gone downhill' in the past couple of decades as political correctness has become prominent, fail to recognize that colleges have always suffered from institutional weaknesses. Calls for reform usually focus on symptoms of the institutional design, not the roots of the institution that effectively grant strong property rights over control of resources to faculty employees.

One of the most commonly criticized aspects of higher education is the institution of tenure. The criticism fails to recognize that tenure is not *the* problem, the lack of for-profit status of colleges severely limits how effectively the organizations can be operated. Contrary to popular beliefs, we argue, to illustrate the long-term consequences of institutional design, that tenure was an effort by competent professors to force quality on colleges. The fact that tenure is 'abused' in practice is because of the incentives of the trustees and administrators of colleges, not a legal constraint that prevents dismissal of incompetent faculty members. There is nothing wrong with tenure; the issue is the design of universities as described above. Tenure is a prominent feature that tends to draw attention because it appears to be unique.

Tenure is not new

Tenure as we know it today in the United States was initialized in the early 1900s, but before that colleges did not dismiss professors willy-nilly. By 1820, Harvard appointed professors with 'indefinite' terms. Other colleges operated the same way. As faculty ranks were created, faculty members were given the opportunity to work for advancement. However, if a faculty member was not promoted, it did not mean that he might not remain indefinitely at the college at the same rank. The employment of faculty was formally insecure because most college charters stated that faculty held their positions 'at the pleasure of the trustees'. In practice, faculty had a reasonable expectation that their tenure would continue.

It was not until the twentieth century that the faculty appointment system evolved to one that requires a regular faculty member to leave a university if not promoted or tenured after a certain number of years (Metzger 1973, p. 122). Before tenure became formal, most faculty appointments were year to year, but it was rare for a college not to reappoint all members of the faculty. A survey of 22 major universities in 1910, before tenure was adopted, shows that only faculty members at the rank of instructor were reappraised annually. Those in the professorial ranks were said to hold their positions with 'presumptive permanence'. The rules varied for assistant professors; at approximately one-third of the colleges surveyed, such appointments were considered permanent; at most colleges, it was a multiyear appointment subject to renewal. 'In all cases the meaning is the same, that the appointment is for life to the age of retirement, provided the appointee is efficient. … Appointments of professors and associate professors are practically permanent' (van Hise 1910, pp. 58–9). Another professor observed, 'In practically all of the larger institutions professors enjoy indefinite or permanent tenure upon the first appointment' (Sanderson 1914, p. 892).

A desire for higher standards

The American Association of University Professors (AAUP) was formed in 1915 by faculty in secure positions at leading universities who suffered no employment threat from the lack of formal tenure; they lived under the system of presumptive permanence. As distinguished scholars, all could easily find comparable positions at other colleges. The AAUP founders proposed that it undertake 'the gradual formation of general principles respecting the tenure of the professorial office and the legitimate ground for the dismissal of professors' (Metzger 1973, pp. 135–6). A key part of the process was the use of faculty committees to review faculty appointments, rather than leaving such decisions completely at the discretion of the higher administration.

Even before the formation of the AAUP and the acceptance of formal tenure by colleges, faculty participation in the recruitment and selection of colleagues had increased, and administrators were consulting more often with the faculty on many areas of university life. During the late 1800s and early 1900s, most large universities created faculty–administrator committees to determine many university policies. The adoption of this practice was forced by the increasing size of universities, which made it difficult for administrators to monitor every aspect of the organization. The larger size of universities also reflected the increasing sophistication of academic disciplines, which then required specialization to allow knowledgeable decisions about faculty appointments and curriculum.

Trustees were passive about faculty employment decisions. By 1910, all but one of the leading universities surveyed 'reported that their governing boards

simply ratified the president's nominees for faculty positions' (ibid., pp. 142–3). The point is that before formal tenure as we know it was established, the faculty appointment system operated much as it does today. Trustees did not run the institutions. Hence, one leading academic, writing in 1918, could say that 'governing boards – trustees, regents, curators, fellows, whatever their style and title – are an aimless survival from the days of clerical rule, when they were presumably of some effect in enforcing conformity to orthodox opinions and observances, among the academic staff' (Veblen [1918] 1957, p. 48).

In 1915, the newly formed AAUP appointed a committee that recommended tenure rules similar to the modern form of tenure. The committee recommended that tenure be granted after a ten-year probationary period, during which time an assistant professor could be dismissed. At the end of the probationary period, the faculty member would be dismissed or granted tenure (Report of the Committee 1932, pp. 390–91). Given the AAUP leaders' incentives at that time and the tenor of their discussion on the record, it is clear that they believed a tenure track would avoid the problems created by the prevailing 'presumptive permanence' rule then in place. That is, leading scholars wanted to be sure they did not have to tolerate non-productive colleagues imposed on them by administrators who did not know or care about quality within a given discipline. The adoption of tenure tracks was intended to *raise* the standards for faculty.

> If this profession should prove itself unwilling to purge its ranks of the incompetent and the unworthy, or to prevent the freedom which it claims in the name of science from being used as a shelter for inefficiency, for superficiality, or for uncritical and intemperate partisanship, it is certain that the task will be performed by others … who lack certain essential qualification for performing it. (Committee Report on Academic Freedom 1915, p. 34)

Colleges adopted the AAUP's recommendation. A report in 1924 found that in some colleges it was still the case that 'after the first year's service a man is practically a fixture unless something very unforeseen happens' (Brooks 1924, p. 498). At schools with tenure tracks for assistant professors, most had a two- or three-year track, not a ten-year track as recommended. 'Reappointment, especially if made more than once, carries with it a strong presumption of permanence' (ibid.). A report in 1932 found that 'There is a presumption of permanency for assistant professors in 91 [of 283] of the institutions studied' (Study of Tenure of University and College Teachers 1932, p. 256). By that time, approximately half of the colleges had adopted formal tenure rules, including provisions for dismissal in case of improper behavior or incompetence.

In 1940, the AAUP recommended that the probationary period be reduced to seven years, which is still the usual standard. Colleges that subscribe to

AAUP standards, as most do, must make tenure decisions before the probationary period ends, or tenure is presumed to have been granted. The 1958 'Statement on Procedural Standards in Faculty Dismissal Proceedings' indicates that dismissal of tenured faculty 'will be a rare exception, caused by individual human weakness' (quoted in Joughlin 1969, p. 41).

Tenure Today

Most college faculty rules now state something to the effect that to keep tenure one must maintain the standards of the profession. That is, one must continue to be a decent teacher of competent material and maintain evidence of scholarly ability in one's areas of academic expertise. The rules are basically the same at most public and private universities. In other words, there is no legal protection for faculty members who stop developing intellectually, do not meet the standards of their discipline, or become unprofessional in the classroom. This point is worth repeating: tenure does not protect faculty who become incompetent.

Each college, under the direction of its trustees, establishes guidelines for earning and retaining tenure (McHugh 1973, p. 255). The courts require universities to establish relatively formal procedures for handling the dismissal of faculty. Court decisions in this regard are like those in other areas of employment law: if employees have been promised certain procedural safeguards, those procedures must be followed. State employees, including college faculty, are due certain constitutional protections against political retaliation. The courts do not view tenure as a lifetime sinecure not dependent on performance. As long as a college follows its procedures properly, it is free to establish whatever competency standards it wishes for its faculty and to enforce those standards.

It is often asserted that tenure creates a right or a property interest in the employment position, but as the US Supreme Court observed, a professor's claim to an entitlement in a faculty position at a public college must rest on more than his or her mere 'subjective "expectancy"' of continued employment (*Perry v. Sindermann* 1972, p. 603). For private institutions, the claim to particular procedural rights prior to discharge must rest on a contractual relationship between the employer college and the professor. Such contractually derived rights may be written into a professor's employment contract, or the contract may incorporate, by reference, a statement of the institution's tenure policy as it appears in a faculty handbook, in published policy statements by the governing board, or in the institution's bylaws. Written statements of an institution's tenure policy are generally viewed as an implied term of faculty employment contracts even if the contracts contain no specific reference to the policy. This is true also at public universities because contractual rights to tenure are a relevant part of the employment process.

Why does tenure appear, to many critics and even to many who hold tenure, to produce undesirable results? The idea of tenure not being granted until after a probationary period of seven years was a quality-enhancing move that was an improvement on the old system of presumptive permanence. It was pushed for by leading professors who had reason to be concerned about the quality of their colleagues, not their own job security. But how tenure has come to work in practice is a reflection of the effective property rights regimes that dominate universities, which includes the lack of incentives for trustees and administrators to impose meaningful performance reviews on faculty.

Tenure legally does not grant faculty rights to behave in an improper or unproductive manner. Tenured faculty who become incompetent, fail to perform their duties, or behave in a grossly improper manner can be fired. A survey of appeals court cases indicates that does not happen often (Amacher and Meiners 2003). Given the hundreds of thousands of tenured faculty members, if there were many dismissals, there would be likely to be more cases, and such instances would be more common. The focus of discussions regarding problems in higher education should not be on the institution of tenure, but rather on the structure of higher education and the implicit property rights created within the peculiar institutions we know as universities.

Production of learning

Despite the incentive problems that afflict institutions of higher education, teachers teach and do research; learning does occur. Compared to the production of widgets, where a firm produces a clearly defined product and sells that product to a consumer, who obtains clear rights to the product, and guarantees as to its qualities, the purchase of educational services is a more complex transaction.

Learning requires joint inputs of students and teachers. These inputs cannot be easily measured or monitored. Such jointness, and the inability of teachers to claim a share of students' increased earnings due to successful instruction, creates incentive effects. Despite the broad range of choices granted to students at most universities, they do not have equal rights to degrees. There are market forces that influence the demand for various majors and the supply of students who can satisfy the requirements of various degrees. Universities may not duplicate what would be the behavior of for-profit institutions, but they appear to approximate such behavior (see Meiners and Staaf 1995 for greater discussion on this point).

Private, non-profit universities do a better job teaching, on average, than do public universities. Many private colleges advertise small class size. Smaller classes make it less costly for instructors to monitor the performance of each student, make it more costly for students to shirk by daydreaming in class,

make it less costly to assign term papers and essay examinations, and give students more incentive to complain about shirking by faculty. Rhetoric about devotion to quality teaching aside, public colleges are more likely to have large classes that allow more shirking by students and lower expectation by students about assistance from faculty. A small decline in enrollment, or decline in the quality of applicants, is costly for a relatively small private school. In mass-enrollment public universities, changes in reputation are less costly to the administration and to faculty members, so there is less reason to press for quality teaching.

Across disciplines, whether schools are private or public, there are differences in production processes that reflect market forces. For example, the faculty in engineering colleges have high opportunity costs in the market that are reflected in their compensation packages. Faculty in colleges of education and liberal arts tend to have few non-education alternatives. Engineering colleges are notorious for having standards such that students tend to depart for easier programs. Having good employment alternatives means the engineering faculty can demand better performance of their students. Faculty in other colleges might like to require more of their students, but have much more to fear in their own future employment should enrollment in their departments decline. Engineering is a profession that has standards; a college that passes out engineering degrees to poorly trained students will soon be known to prospective employers. Poorly trained students result in a decline in the value of the engineering faculty members' reputations and their market alternatives. Hence, the engineering faculty have more incentives to focus on quality of student output. In colleges of education, at the other extreme, there is little agreement about what constitutes good outcomes, so discussion tends to focus on measures of inputs as signals of quality. No one seems fooled by these self-serving discussions.

Caveat emptor: assignment of liability

There would be no demand for education if students knew what they were demanding. Education is an extreme experience good; students do not fully appreciate the good or the cost without large investments of time (Nelson 1974). Universities warrant very little; even for-profit schools do not give unconditional warranties. For example, for-profit flight schools, where one trains to be a pilot, do not guarantee that one will become a licensed pilot simply by paying tuition fees and attending school. Learning is a joint activity that requires input from the students as well as the instructors. Unconditional guarantees would expose schools to opportunistic behavior by students.

The nature of education is such that it is one of the few areas of law in which *caveat emptor* dominates. Whether produced by for-profit or non-profit organizations, the jointness of the production of learning means that the

rule of *caveat emptor* is efficient as it has evolved over centuries of experience. Imagine if the law were to impose liability on universities for the failure of students to learn or be successful in desired careers. With no ability to contract around the rule, as required under the Coase theorem for efficient results to emerge, or for universities to hold students in slavery or indentured servitude to help limit opportunistic behavior by students and capture financial returns to education, assignment of liability to colleges is not possible. Legislatures and government agencies persist in demanding evidence of quality in public universities, but, given the extreme difficulty of measuring quality of output, this is a fruitless task, especially when politically motivated, and is likely to lead to perverse actions by college administrators who seek to meet artificial goals.

Commons in college clubs

While most of the value in education is captured by students who obtain it, many university resources are common rights. This is not true of housing and food service, where the benefits are mostly private and there are good private alternatives, but it is the case with many of the educational resources. For example, library books could be rented to students so that the high demanders pay more and seats in the classes of the most popular professors could be auctioned for a tuition premium, but we do not observe such actions at private or public universities. Part of the reason could be the transaction costs, but computers have driven such costs to trivial levels.

The absence of more extensive price discrimination and charging for use of each university service may be to help instill loyalty and sense of belonging to a club (Buchanan 1965). Colleges, like private golf clubs, often discriminate on criteria other than price. Harvard could charge much higher tuition than it does, but, like an exclusive club, it wishes to have a list of applicants to draw upon to ensure certain characteristics of average students. Given that employers often know little about the relative merits of individual college graduates, they use the name of the school as valuable information. Schools have reputations that provide information to prospective employers. To ensure that this reputation is maintained, colleges have strong incentives not only to discriminate on entry on criteria other than price, but to allow students to have common property rights to school resources, to encourage the production of graduates who share common experiences that help increase long-term loyalty to the club.

Conclusion

Colleges in the United States have seen productivity improvements in many areas. Food services have improved as private firms have taken over from college-run kitchens; college-run dorms have been replaced by privately-

operated living quarters; computers have allowed registration lines to be eliminated; libraries are hooking into computer databases; but academic departments are run much as they were a century ago. Unless there are changes in the structure of universities there is little reason to expect efficiency improvements.

University administrators can only attempt to ape efficiency techniques they observe in the private sector, much as central planners in communist countries attempted to obtain prices and other valuable information from market organizations. Incentives to improve universities, public or non-profit private, are severely constrained by lack of strong financial market signals that for-profit firms enjoy. However, market forces are at work at all times and in all places. When we consider universities, which do compete with each other, and where there is internal competition for resources, as places where rights are exchanged and enforced, we can understand that many of the common features of universities are to be expected given the constraints faced by the participants, students and providers, in that particular market.

Private universities appear to do a better job at teaching. Whether their focus is graduate or undergraduate education, private schools dominate national rankings of quality institutions. They do this despite the benefit of state dollars being the basis of their budget. State universities, as bureaucracies, naturally focus on how to please political decision makers, rather than focus more on attracting better-quality paying students, and obtaining donations from alumni and other donors that wish to support their educational mission. Both state and private colleges appear to suffer from relatively weak leadership due to their ownership structure that means a lack of external market financial discipline. In both sets of institutions, faculty especially can gain effective control of significant resources to help ensure that their positions will not be eliminated as there are shifts in the demand for different mixes of educational output. Nevertheless, private schools must still make it in the higher education market and remain more focused on that market than on pleasing political masters. As is the case in any area of economic activity, the design of the institution plays a critical role in determining the incentives of those who work in the area and how they will devote their efforts.

References

Amacher, Ryan C. and Roger Meiners (2004), *Faulty Towers: Tenure and the Structure of Higher Education*, Oakland, CA: Independent Institute.

Brooks, R.C. (1924), 'Tenure in colleges and universities', *School and Society*, **19** (487), 26 April, 497–501.

Buchanan, J.M. (1965), 'An economic theory of clubs', *Economica*, **32**, February, 1–14.

Committee Report on Academic Freedom (1915), *Bulletin of the American Association of University Professors*, December, 15–43.

Demsetz, H. (1967), 'Toward a theory of property rights', *American Economic Review*, **57**, May, 347–59.

Honan, W. (1994), 'New pressures on the university', *New York Times*, 9 January, sec. 4A.
Joughlin, Louis (ed.) (1969), *Academic Freedom and Tenure: A Handbook of the American Association of University Professors*, Madison, WI: University of Wisconsin Press.
McCormick, R. and R. Meiners (1988), 'University governance: a property rights perspective', *Journal of Law and Economics*, October, 777–95.
McHugh, William F. (1973), 'Faculty unionism?', in Bardwell L. Smith (ed.), *The Tenure Debate*, San Francisco, CA: Jossey-Bass, pp. 244–61.
Meiners, Roger (1995), 'The history of American higher education', in John Sommer (ed.), *The Academy in Crisis: The Political Economy of Higher Education*, New Brunswick, NJ: Transaction, pp. 21–43.
Meiners, Roger and Robert Staaf (1995), 'Property rights in academe', in John Sommer (ed.), *The Academy in Crisis: The Political Economy of Higher Education*, New Brunswick, NJ: Transaction, pp. 197–216.
Metzger, Walter P. (1973), 'Academic tenure in America: a historical essay', in *Faculty Tenure: A Report and Recommendations by the Commission on Academic Tenure in Higher Education*, New York: Jossey-Bass, 93–159.
Nelson, Ph. (1974), 'Advertising as Information', *Journal of Political Economy*, **82**, July/August, 729–54.
Paarlberg, D. (1992), 'The land grant college system in transition', *Choices*, **7** (3), 45.
Perry v. Sindermann (1972), 408 U.S. 593, 92 S.Ct. 2694 (Sup. Ct.).
Report of the Committee of 1915 (1932), *Bulletin of the American Association of University Professors*, May, 378–91.
Rosovsky, Henry (1990), *The University: An Owner's Manual*, New York: Norton.
Sanderson, E.D. (1914), 'Definiteness of appointment and tenure', *Science*, **39**, 19 June, 890–96.
Study of Tenure of University and College Teachers (1932), *Bulletin of the American Association of University Professors*, April, 255–7.
Van Hise, Ch. (1910), 'The appointment and tenure of university professors', *Journal of Proceedings and Addresses of the Association of American Universities*, **12**, 50–61.
Veblen, Thorstein (1918), *The Higher Learning in America*, reprinted (1957) in New York: Hill & Wang.

Index