

INSTITUTIONALISING
COMMITMENT AND CREDIT

Historical Tests of a Microeconomic
Theory of Innovation, Governance
and Growth

by

Simon John Morley

London School of Economics and
Political Science

University of London

A thesis submitted in partial fulfilment
of the requirements for the degree of

Doctor of Philosophy

2004

UMI Number: U615472

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U615472

Published by ProQuest LLC 2014. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

THESES

F

8262

978116



ABSTRACT

This study investigates a factor that observation suggests is an important causal factor in economic outcomes, namely governance. Because contemporary economic theory provides little explanation of the observed link between innovation, governance and growth, a new theory of microeconomic activity is proposed; one that brings together the efficient markets hypothesis of financial economics, and the role of property rights that is central to the new institutional economics of North and Williamson. In particular, a conceptual model is set out, suggesting that the existence of governance mechanisms, which enable credible commitments to be given regarding future economic performance, is a crucial enabling factor for economic growth; because, within a market economy with imperfect information, they provide the right incentives for the cooperation necessary for socially beneficial innovation. The central body of the thesis seeks to test this model by applying it to four separate but important historical epochs, using secondary historical sources. Ancient and medieval Europe provides the first two applications. Following this, detailed consideration is given to England around the eighteenth century, and then to the United States in the late nineteenth and early twentieth centuries.

The development of commitment mechanisms, first in the form of legal contracting, and then in public accounting, is described, and, when this also enables the growth of financial markets, economic growth in the form envisaged by the theory can be observed. Furthermore, the thesis argues for the scientific value of this commitment-credit theory, because of the observability, and so testability, of the concepts at its core. This suggests that, from the point of view of managers and government policymakers, this is an improvement over many other economic theories, and the neoclassical theory in particular, which builds on introspectively derived axioms, and rushes to give policy recommendations, without first testing the validity of the theory.

TABLE OF CONTENTS

Abstract	2
Acknowledgments	6
1. Motivation, Model and Methodology	7
1.1 Observations on Governance	8
1.1.1 Specific Thesis: Governance is Key	9
1.1.2 Practical Consequences of Theoretical Failure	10
1.1.3 Initial Observations of Practical Failures	14
1.1.4 Initial Observations of Practical Success	18
1.2 Commitment-Credit Theory	19
1.2.1 Logic of Modern Finance as a Starting Point	21
1.2.2 Legal and Financial Logic	24
1.2.3 Economic Consequences	28
1.2.4 Scientific Tests	31
1.2.5 Presentation of Results	36
2. Historical Background	39
2.1 Ancient Greece and Rome	40
2.1.1 Commitment	40
2.1.2 Credit	47
2.1.3 Consequences	52
2.2 Medieval Europe	57
2.2.1 Commitment	60
2.2.2 Credit	64
2.2.3 Consequences	69
3. Industrial Revolution in England 1700 - 1850	77
3.1 Commitment	85
3.1.1 Legal Environment	85
3.1.2 Rise of <i>Assumpsit</i>	92
3.1.3 Legal Organisation	98
3.1.4 Bills and Notes	100
3.1.5 Parliamentary Innovations	101
3.1.6 Legal Revolution	104
3.2 Credit	105
3.2.1 Public Finances	106
3.2.2 Private Working Capital	110
3.2.3 Private Investment Capital	115
3.2.4 Capital and Creative Destruction	117
3.2.5 Financial Revolution	117
3.3 Consequences	118
3.3.1 Output Growth	118
3.3.2 Firm Growth	122
3.3.3 Financial Structure	125
3.3.4 Business Cycle	127

3.3.5 Boulton and Watt.....	129
3.3.6 Conclusions.....	134
4. Managerial Revolution in the USA 1850-1950.....	135
4.1 Commitment	137
4.1.1 From Bookkeeping to Accounting.....	138
4.1.2 Accounting for Management Control.....	140
4.1.3 Railway Accounting.....	143
4.1.4 Financial Accounting.....	146
4.1.5 Managerial Accountability.....	150
4.1.6 Accounting Revolution.....	153
4.2 Credit	154
4.2.1 Continental Scale.....	155
4.2.2 Rise of New York.....	156
4.2.3 Civil War Interlude.....	159
4.2.4 Railroad Dynamics.....	160
4.2.5 The Birth of Corporate Finance.....	162
4.2.6 Self Funded Consumer Business Expansion.....	163
4.2.7 Competitive Investment Banking.....	166
4.2.8 Corporate Financial Revolution.....	170
4.3 Consequences.....	170
4.3.1 Output Growth.....	171
4.3.2 Form Elaboration.....	174
4.3.3 Capital Allocation.....	177
4.3.4 Product Cycle.....	180
4.3.5 General Motors.....	182
4.3.6 Conclusions.....	188
5. Governance in Economics.....	189
5.1 New Institutional Economics	190
5.1.1 North and Thomas.....	190
5.1.2 Coase.....	195
5.1.3 Williamson.....	196
5.1.4 North: the Sequel.....	200
5.1.5 New Institutional Economics and Observability.....	203
5.2 Keynesian Economics	206
5.2.1 Theoretical Background to Keynes.....	207
5.2.2 Historical Background to Keynes.....	210
5.2.3 Keynes' Theory.....	212
5.2.4 Keynes' Impact.....	216
5.2.5 Keynesian Economics and Observability.....	220
6. Theory in Practice.....	222
6.1 Scientific Method.....	223
6.1.1 Underlying Thesis: Science is Observation before Formalisation.....	224
6.1.2 Popper's Scientific Logic of Falsification.....	226
6.1.3 Neoclassical Falsification.....	229
6.1.4 Commitment-Credit Model Falsifiability.....	232
6.1.5 Distinguishing Historicism and Historical Data.....	236
6.2 Social Science.....	240
6.2.1 Kuhn's Metatheory.....	241
6.2.2 Neoclassical Economic Theory as Kuhnian Science.....	246
6.2.3 Neoclassical Alternatives as Kuhnian Science.....	248

6.2.4 Commitment-Credit Theory as Kuhnian Science.....	252
6.2.5 Scientific Contribution of the Commitment-Credit Model.....	255
6.2.6 Critique of Scientific Distinctiveness.....	260
6.2.7 Practical Contribution of Commitment-Credit Theory.....	263
7. Bibliography	269

ACKNOWLEDGMENTS

I want to thank everyone who has helped with this project. Although it has been an almost wholly solitary effort, as is perhaps reflected in its idiosyncratic nature, many people have helped sustain the motivation that has been the most important resource in its completion. Staff and students at the IIM obviously played an important role, and I would like to mention in particular Prof. Peter Abell, Prof. David de Meza and Dr. Tobias Kretschmer. Barclays Bank, and in particular my Premier Account Manager Ian Astill, probably without knowing it, provided critical, albeit temporary, financial support in allowing it to progress. Along the way I earned an income, including in the most recent cases accommodation, working for the LSE Information Systems Department, Groupe HEC in France and the NGO Civic Education Program at Odessa State Economics University. I would like to apologise for the problems that were an almost inevitable result of trying to combine completing a self-funded PhD with other responsibilities, and thank my colleagues who, particularly in Odessa, welcomed me as a part of their team, and helped alleviate the worst of the isolation that this type of project inevitably involves. I am grateful to Amazon, in the UK, France and US, including in some particularly useful cases their second hand book service, who provided most of the raw material on which this study is based. I would like to say 'спасибо огромный' to Anya, who made the last eighteen months of this project, which could have been the worst part, into the best of times. Most of all, as always, I thank my parents, whose ever supportive attitude was probably the only constant element throughout this long project, which has dominated my life for more than half a decade.

Simon Morley,
Westcliff, Essex,
June 2004.

Chapter One

MOTIVATION, MODEL AND METHODOLOGY

On 2 December 2001 Enron, once the seventh largest company in the United States, filed for bankruptcy protection. Then, on 2 January 2002 Argentina, once almost as wealthy as the United States, formally defaulted on its debt. It is unlikely that there was any direct connection, except perhaps a global economic downturn and the psychological significance of the end of the year; but the coincidence in timing helps in suggesting a similarity in the functioning of both nation states and large corporations. Argentina and Enron both failed because they were unable to fulfil commitments they had previously given in return for financial credit.

The consequences for both were drastic. In the run up to formal default the government of Argentina changed hands five times in two weeks, and GDP declined by the annual equivalent of 16.3 percent over just three months (Cibils, Weisbrot et al. 2002). The impacts on many citizens' lives cannot be easily described in a few words. Although many parts of Enron survived under new parents, thousands of people lost their jobs; and many employees who had invested their pension savings in the company's shares found their material prospects in retirement transformed overnight from comfortable to struggling. The Enron fallout even brought about the end of its firm of auditors, Arthur Anderson, previously a leading global firm.

Although this study is not directly about Argentina or Enron, not least because it was well towards completion when these events occurred, it does go some way to explaining what happened to them. The commitment and credit mechanisms that failed for Argentina and Enron pervade our world, and to a large extent make our economic world what it is. The mechanisms have brought, and continue to bring, massive benefits to billions of people on

our planet; but as these cases show they are not without their downsides. However, before being able to deal effectively with many of the problems that accompany the system, it is first necessary to understand the logic that underpins the normal functioning of the system. That is what this study aims to do.

1.1 Observations on Governance

The focus of this study is to investigate an important causal factor in economic outcomes, namely the role of governance mechanisms. These provide commitment mechanisms, both directly and indirectly through financial markets. Without them the innovation that underlies economic growth is unable to be properly directed. This is obviously a big subject, but perhaps more importantly, one that will have been extensively studied already. Or at least one would assume that fundamental questions such as these would have received considerable attention, from the economics profession in particular over the hundred or so years of its existence. Yet if it has, it has not survived into the contemporary literature. Some consideration has been given to governance, and this will be reviewed, but with one exception, it is not linked with innovation. The exception is the work of economic historians North and Thomas, whose analysis of the institutional changes leading to the industrial revolution in England has many similarities to this thesis, but also important differences. In many ways this thesis could be considered a work of economic history, but its motivation is to provide a new economic theory that offers to be more useful than many already in existence, at least from the point of view of managers and government policymakers. So its contribution is assessed not in terms of economic history, or contemporary economic theory, but as a scientific theory, the confirming historical tests of which justify its value to practitioners.

In many ways the objective of the project of which this thesis is the final product has been to provide a bridge between management and economics,

although it has only been able to do this by suggesting that major changes are needed in the content and process of microeconomic research. However, this is not primarily a critical work, it is first of all constructive. Clearly, if this thesis does identify a fruitful new direction for economic research then much more work will be needed. What is sketched out here is a logical framework and its application, in short a model and a methodology, that together make up a new theoretical approach, albeit one that draws on extensive previous work across the social sciences.

1.1.1 Specific Thesis: Governance is Key

The specific thesis at the heart of this study is that particular laws and regulations governing economic interactions largely determine economic outcomes, in much the same way that chemical interactions are responsible for biological outcomes; and furthermore, that these economic phenomena can only be well understood once the logic that links the two levels has been worked out. These legal mechanisms commit economic actors to future actions in a way that is mutually beneficial to the actors involved, and ultimately to the society in which they live.

What is proposed here is that the ‘throwaway lines’ as to the importance of property rights and a functioning legal environment of many economists, before they proceed to focus a great deal of attention on formalising agent decision rules, puts the cart before the horse. It is becoming a commonplace that governance and ‘the rule of law’ matters; particularly given the problems caused by placing too much reliance on the neoliberal ideologues who failed to take this into account in trying to manage post-communist reform in Eastern Europe and structural adjustment in sub-Saharan Africa. This study goes one stage further. It argues that it is the governance mechanisms that are of first order importance and, given those constraints, agents then have a wide variety of decision strategies open to them, although their actual choices are not necessarily as important to the overall economic outcome as the attention given them by conventional methods of economic analysis would suggest.

Although economists who claim that economic analysis is marginal analysis would obviously disagree, the focus here on systemic outcomes of interactions between large numbers of independent agents places this theory firmly in the domain of economic enquiry. The second half of this chapter specifies exactly what is proposed. It sets up a model that encapsulates this logic, and describes, in an abstract way, how governance connects economic interactions and economic outcomes.

This project also differs from conventional economic theory because it suggests that correspondence with observational evidence is more important than mathematical language in providing a scientific explanation of both economic success and economic failure. The greater part of this study uses historical evidence as a ‘natural experiment’ to attempt, in a Popperian sense, to falsify the model. For Popper, and even more so for later intellectual historians, science is a communal activity, so falsification is best undertaken by people other than those who propose the theories. Nevertheless, although this falsification is done in a somewhat contrived way, it does provide a linking historical narrative through which the explanatory power of the model can be demonstrated, and allows the basic elements of the model to be further elucidated.

Enron and Argentina provide recent examples of devastating economic failure, and in doing so also highlight the danger that unscientific economic theorising is not just doomed to irrelevance, but can also damage the lives of huge numbers of people. Before moving on to outline the new theory, it is worth clearing the ground a little more by briefly considering some of the evidence that is already available on the role of conventional economics and economists in those disasters.

1.1.2 Practical Consequences of Theoretical Failure

Enron and Argentina are both problematic economic experiences, and for two reasons above all, and the reasons are not as unrelated as they might

appear at first sight. First of all, they are problematic because those who lost out most severely were those least able to take action themselves to recover, namely pensioners, because most of their more productive years were over. Secondly, they are problematic because, in both cases, until shortly before their abrupt failures, they were held up as paradigmatic poster children of neoliberal economic policies.

In the case of Enron, right from its inception, the funds set aside to pay the pensions of people employed by the firms that had merged to make up Enron were put at risk, essentially to protect the jobs of the senior management. Probably more serious was the holding of very high proportions of Enron stock in many employees' personal pension funds. While diversifying a portfolio is the first rule of rational investing, particularly given efficient capital markets, there were two reasons why many employees failed to adopt this approach. First of all, the company encouraged employees to spend their personal pension plan contributions on its own stock by matching purchases, and secondly, many people wanted to believe that they would benefit in the same way as employees at Microsoft, many of whom were millionaires as a result of the Microsoft stock they had been given by the company. How were Enron's employees to understand the difference between their company and one like Microsoft that created real value? But when Enron crashed, so did the pension funds that were largely made up of its stock (Fusaro and Miller 2002, 122).

In Argentina, one of the reforms undertaken in the public sector was to privatise pension contributions, by making these compulsory for new workers and providing incentives for existing contributors to switch. This clearly created a massive hole in the existing 'pay as you go' system, where the operative principle is that current workers pay the pensions of retired workers through their taxes, because current workers were now only paying for their own pensions. The situation should not have been as serious as it might have seemed, because the private funds were required to hold substantial

proportions of their debt in, supposedly low-risk, government debt. The money raised from this would be used to fund existing pension obligations; just leaving current revenues to pay the interest and gradually contribute to what was essentially a one time capital charge (Lewis 2002, 163, 164). However, reality caught up when public sector pensions were one major area to be cut in the emergency reforms of July 2001. The private sector did not escape however, as pension funds were forced by the government to increase their holding of government debt, the value of which fell precipitously as default increasingly became a likelihood, and then a reality, so decimating the value of these funds in turn (Lewis 2002, 173, 174).

Enron rode the crest of two popular waves before it crashed so spectacularly, that for energy deregulation and then the 'new economy' of internet based businesses. Its CEO for most of its existence was Ken Lay, and Fusaro and Miller point out that he was well educated, but that unlike most chief executives who have at least some education in engineering or management, all Lay's degrees were in economics. They argue that what drove Lay and Enron was not any fundamental innovation, which is what normally lies behind any firm that grows so quickly, but simply an idea, that of free markets (Fusaro and Miller 2002, 2). Lay's successor as CEO, until his retirement under dubious circumstances and Lay's return, was Jeffrey Skilling. Apparently, while CEO he replaced his vanity number plate, which read WLEC for world largest energy company, with one reading WMM, for we make markets. His message was that having become the leading gas and electricity trading company, Enron would use the internet to create markets in a whole host of as yet unimagined commodities. Of course, for Skilling these markets were not to be public goods, but cash cows to service Enron's ever-growing debts (Fusaro and Miller 2002, 70).

Lewis describes Argentina as the most assiduous disciple of the International Monetary Fund (Lewis 2002, 174). And, in retrospect, lending from the IMF was a major factor in preventing an earlier recognition of the deep structural

problems in the economy, as well as possibly making the consequences worse when the problems did finally surface. Much as with Enron, Argentina bet its future on the neoliberal consensus of 'free' markets, hoping that the growth that would naturally follow over time would generate the wealth necessary to tackle the other issues. The problem was that the Argentine economy was not growing. And privatisation, starving the public sector of funds, and imposing strict monetary discipline, did not come close to tackling the underlying problems of insufficient public revenue and entrenched special interests.

It is the connections between pensions and free markets that highlight the need for an economic science that the neoclassical model fails to provide. In effect, neoliberalism is the policy framework that is the natural consequence of the neoclassical model. In the same way that equilibrium is a static concept and, particularly with neoclassical Arrow-Debreu general equilibrium models, all future states can apparently be translated into currently tradable commodities, so neoliberalism prescribes markets now, to create the best possible future, with no need to build any institutional infrastructure. Yet this approach was demonstrably wrong for Enron's and Argentina's pensioners. While neoliberalism emphasises individual responsibility, and individual rewards, pensioners do not live in a static world. They have already made a large proportion of their contribution to their society and economy, and have little left to bargain with when the shiny new market is introduced. While markets are necessary for sustainable economic growth, they are clearly not sufficient.

Moreover, it is not the role of economic science to tell managers and policymakers what to do, although it can explain interdependencies and, as a successful paradigm develops, it will be better able to specify more completely and accurately the likely consequences of change in one factor for others. Managers and policymakers will always have to take the responsibility themselves, for success and failure, particularly if they make decisions that the best current theories suggest will have problematic consequences.

1.1.3 Initial Observations of Practical Failures

So much for justifying and motivating the work of this project from past failures, let us now turn to the positive task of setting up an alternative. Before setting out the model in the next section, some initial observations based on Enron and Argentina will provide several issues to consider in the interaction between innovation, governance and growth.

Enron was undoubtedly an innovative firm. It was selected by Fortune magazine as the most innovative firm in the United States every year from 1996 to 2001. In 2000 its revenues made it the seventh largest firm in the Fortune 500 (Fusaro and Miller 2002, 75, 173). Unfortunately, much of that innovation was misdirected, and large parts of the firm subsequently proved to be formidable financial shells, with little underneath the surface.

Enron was created in 1985 from two natural gas companies, Houston Natural Gas and InterNorth (Fusaro and Miller 2002, 4). Right from the start much energy was directed at what were non value-adding activities. Shortly after the merger the company's executives felt it necessary to secure their independence by buying back, at a considerable premium, a substantial shareholding from someone thought likely to attempt to take control of the company, and to do this used 'excess funds' from the employee pension scheme, together with high interest debt (Fusaro and Miller 2002, 5). To both pay off these debts and create the appearance of earnings growth a programme of asset sales commenced (Fusaro and Miller 2002, 12). Because these assets had appreciated in value since they were built or purchased, but were recorded in the accounts at their original cost, as conventional accounting principles dictate, the additional income they raised when they were sold was booked as profit. Although in itself legitimate, this was clearly not a sustainable process, never mind a genuine growth strategy. There were early signs of problems regarding managerial accountability in the late 1980s, when a subsidiary based in New York, away from Enron's Houston headquarters, was able to run up losses of around \$100 million, which was

discovered not by Enron itself, but by regulators (Fusaro and Miller 2002, 21). Unfortunately, both practices continued, and indeed reinforced one another, as Enron reinvented itself from a company that supplied natural gas into one that traded it (Fusaro and Miller 2002, 37).

While energy deregulation and the technology of derivatives trading undoubtedly provided opportunities to create real value, the accountability necessary to create a sustainable business was not regarded as important, so it was not long before Enron's credit ran out. In April 2001 Enron released its results for the first quarter of 2001. At least it released a profit and loss account, which showed its revenue and expenditure. It did not release a balance sheet. When it did release the balance sheet in mid May, its cash reduction amounted to more than it had made in profits. While this was not necessarily a problem in itself, it was a warning signal, one that created uncertainty and was followed by declines in Enron's share price. By the time it released its next quarterly results its earnings were forty percent up, but its revenues declined slightly, while its balance sheet showed its cash holding had gone down by over a billion dollars in three months, almost twice as much as earnings had increased. The Chairman and CEO resigned in what are still mysterious circumstances. Then with the release of its third quarter results on October 16 it seemed that perhaps the previous Chairman, who had returned from retirement, was putting the record straight, by announcing a loss of \$618 million. However, this threatened to reduce the risk rating of its bonds, which was critical because without a top rating Enron would no longer be able to trade the transactions it facilitated, which were all effectively guaranteed by the firm. It officially corrected its financial statements for the previous four and a half years, as a part of negotiations with another firm to buy up Enron as a going concern. But when those negotiations broke down on November 28, there was no choice but to sue for bankruptcy, on December 2 (Fusaro and Miller 2002, 109, 112, 113, 116, 121, 122).

Argentina is a country rich in natural resources, particularly in its educated population and fertile agricultural land. In the 1880s its Gross Domestic Product (GDP) per capita was almost at the level of the United States. Notwithstanding the Barings crisis of 1889/90, the subsequent century shows a mixed record of economic growth, although ultimately growth became not just relatively but absolutely negative in the 1980s and 90s (Lewis 2002, 88-89).

Bouts of very high inflation in the 1980s were brought under control by a series of measures designed to achieve monetary stability, and for a decade they achieved their aim. In January 1990 short-term high-interest deposits were temporarily frozen. Long-term deposits were converted into ten-year government bonds, denominated in US dollars. Fiscal policy was tightened through reducing the number of people employed in the public sector, and also benefited from privatisations. Then in January 1992 a new currency, the Peso, was introduced at a conversion rate of one to the US Dollar, guaranteed against reserves held by the central bank equal to the value of Pesos in circulation (Lewis 2002, 161-163). From 1991 to 1994 the plan succeeded in its objectives. Output grew at just under eight percent per annum and consumer price index inflation fell from eighty four percent in 1991 to three point nine percent in 1994. Unemployment rose, but remained below double digits. The economy also absorbed the external shock of the 1994 Latin American 'Tequila Crisis.' 1995 saw negative growth, but in the subsequent three years it was positive. Average annual total factor productivity increases were positive from 1991 through to 2001, and by the end of the decade export performance was positive (Lewis 2002, 165).

Unfortunately, there were what turned out to be unsolvable problems in the public sector. The pension, education and health systems remained strongly in deficit, which previously had been largely disguised by high inflation. The tax system was problematic, being largely based on indirect taxes, so was also socially inequitable and pro-cyclical. But most serious, as a percentage of

GDP, Argentina's tax base was only half that of countries such as Brazil, Chile, Mexico and the US. Problems were compounded because, while the central government was largely responsible for collecting revenue, the provinces were responsible for spending it, and public services were further devolved as a part of the reform process. As reform progressed, social problems became more serious. Capital replaced labour in the private sector, and the workers displaced, like those who lost out in the public sector, were often the unskilled and middle managers, who found themselves least able to provide what the reformed economy demanded from its workforce. Worst of all was the cancer of corruption that pervaded the political class, and had become the only way a party could maintain its cohesion in government (Lewis 2002, 164, 167, 168, 175). From an economic perspective, corruption means that the rewards of economic activity primarily derive not from creating economic value, but from retaining the support of individual politicians and government officials, so making growth almost impossible.

The stresses in the economic system finally started to bring it to its knees in March 2001, when over \$5 billion was withdrawn from the banking system. Then between April and mid July a further \$10 billion was taken out of the system. New taxes were introduced and public sector salaries were slashed, although parliamentarians exempted themselves and their immediate employees in congress from the wage cuts. A last desperate measure was the freezing of many bank accounts in December, but very soon afterwards the government effectively collapsed, swearing in five different presidents in two weeks spanning the start of 2002, followed by defaulting on its loan obligations, and the subsequent and continuing economic and political crisis (Lewis 2002, 172, 173, 215).

This brief description of the underlying problems that led to the sudden collapse of Enron and the Argentine economy shows that while in both cases there was innovation, or at least the potential for it, undue attention was placed on trying to manipulate the financial position, without focusing on

building the underlying mechanisms to support the real innovation that creates economic value. The importance attached to keeping financial markets happy at Enron, and to the banking system in Argentina, suggests that these indeed play a critical role, but that they are not sufficient in themselves.

In many ways this approach taken by Enron's managers and Argentina's policymakers is not wholly surprising, for as this study will show, the underlying commitment mechanisms that enable economic growth have, until now, eluded detailed attention as a fundamental part of the wider economic system. In part this is because, as we will see, they have evolved largely incrementally, and in response to particular problems. But it is also probably because the issues involved intersect so many conventional academic disciplines that they have been central to the study of none, no doubt for the very understandable desire to avoid unfamiliar but possibly already claimed territory.

1.1.4 Initial Observations of Practical Success

Although the fragmented nature of existing studies of governance poses a problem, it also offers an opportunity, because considerable evidence should exist in the literature across a range of disciplines. There is a risk that from the perspective of well-developed disciplinary traditions an attempt to synthesise the evidence may appear unsophisticated, but a start has to be made. So, what can be observed of the role of governance in economic success? Broadly, the evidence divides into two categories, macro and micro.

The macro evidence takes the form of statistical evidence that governance mechanisms are a major contributor to growth at the scale of national economies. One paper, part of the output of a recent major World Bank study on social capital (Knack 1999, 2), contains a paragraph that is worth reproducing in full.

“A theme repeatedly emerging from cross-country empirical studies is that the difference between developmental successes and failures is largely a

function of incentives facing wealth-maximizing individuals. In some countries, the structure of incentives steers people primarily toward producing new wealth, while in other countries it is easier to gain wealth by diverting it from others. Social capital determines the relative payoffs of production and predation, 'making' and 'taking.' Where social and legal mechanisms for the efficient resolution of prisoners' dilemma and principal-agent games are weak or absent, the private returns to predation increase while the private returns to production fall."

A major source of micro evidence is the work of Hernando de Soto, and his colleagues at the Institute of Liberty and Democracy, in Lima. His recent book, *The Mystery of Capital* (Soto 2000), sets out extensive evidence from a number of countries, across several continents, that the assets held by people in less developed countries fail to provide the economic leverage that similar assets provide in the developed world. In particular, the officially recognised property owner faces an almost insurmountable wall of bureaucracy, making transfer almost impossible. As well as inhibiting business growth in itself, this situation also effectively stifles the development of commercial banking, at the heart of which is the use of property as surety against loaning money, which can accelerate the pace at which a business is able to develop.

Taken together these studies clearly demonstrate the critical importance to economic activity, and to enabling economic growth, of the economic governance mechanisms in the social environment. An economic logic that enabled these issues to be addressed effectively would need to bring to centre stage issues of governance and growth in a dynamic economy, in a clear and concise conceptual framework. The next section sets out in detail such a logic.

1.2 Commitment-Credit Theory

The following five paragraphs sketch out the model that sits at the core of the theory, which the rest of this section expands on. It explains how funding for investment in innovation works, by providing a mechanism for overcoming the problem of potential investors evaluating opportunities with limited information.

1. Innovation: resources are needed to implement an innovation that will create a profit opportunity. Bringing an innovation to market can be expensive, disruptive and have an uncertain outcome.
2. Investment: consumption needs to be forgone in expectation of greater future consumption. For most projects, the innovator does not have immediate consumption opportunities available in enough quantity to be forgone. However, others may be able to provide credit, given the right incentives.
3. Governance: so the innovator needs to be able to make a credible commitment, some form of contract, to give up assets or future revenue, in other words capital, in return for others forgoing immediate consumption in the expectation of future consumption opportunities, even if the innovation opportunity fails to produce its expected gains. The innovator is in the best position to evaluate the risk/return opportunity prior to committing to a contract, and is subsequently free to make decisions with the best available information at the time, given that the contract is 'incomplete;' in other words, makes no attempt to cover every possible eventuality.
4. Finance: investors cannot judge the likelihood of any particular project's success, so their preference is to invest in a capital market that minimises arbitrage opportunities. That is, a market where no one is able to exploit information someone else does not have available. So, in effect, returns are evenly spread between investors, and are based on the aggregate of innovative profit opportunities realised.
5. Growth: general equilibrium and intertemporal preferences determine the overall level of consumption forgone, or saving, in response to investment opportunities. And, with sufficient information and incentives available to innovators, they will realise profit opportunities

that improve the consumption available in the economy, in other words creating economic growth.

The implications of taking this perspective prove to be deep and profound, throwing light on such apparently diverse phenomena as capital markets and economic development, as the remainder of this section shows. The next subsection identifies modern finance as the logical starting point for this study. The elements of the model are then filled out in more detail. Once this has been done some logical consequences are set out. The central claim of this thesis is that an economic model should be judged on its ability to explain economic experience, and the final two sections of this chapter explain first in principle how this assessment will be done, and then how this relates to the overall structure of the thesis.

1.2.1 Logic of Modern Finance as a Starting Point

Work in the last three or four decades in one area of management research, namely finance, does go some way towards bridging the gap between business practice and economic theory, and in many ways provides a sounder logical platform on which to develop a new microeconomic theory than existing neoclassical microeconomic theory.

The development of finance as a subject of academic study separate from economics, and separate from a primary concern with financial management as a professional activity, can be largely traced back to the observation in the 1930s that stock market prices move in ways that are indistinguishable from random, and random selection is as good a prediction of portfolio performance over time as any other method. Bernstein emphasises that as 1933 was the date that this result was first published, extraordinary computational resources for that time were required for the study; which incorporated over eleven thousand buy, hold or sell recommendations (Bernstein 1993, 33). This observation has stood the test of time, and is now labelled the efficient market hypothesis. In this form it stipulates that, in an

efficient market, all information is taken into account in the price of a financial security. Not surprisingly, numerous elaborations to try to make this observation more precise, and so more amenable to more sophisticated statistical testing, have been made; but the basic idea is easy to grasp, and firmly based on observational evidence.

Modern financial theory is now widely applied by people active in financial markets, far more than the deterministic price theory of neoclassical economics is applied anywhere. The random nature of financial security prices, and their numerical nature, clearly leads to the expression of these ideas in a mathematical form. Ironically, the mathematics used is far more sophisticated than that used in neoclassical economics, and most people working in this area trained not as economists, but as physicists or applied mathematicians.

However, numerical data is only one form of data, and mathematics only one form of logic; and clearly there is a close relationship between the two. The past twenty years have made it clear that scientific knowledge with a high degree of predictability, within well-understood limits, does not have to be expressed mathematically. In particular, extraordinary developments in molecular biology rely on the logic of chemical reactions in determining biological outcomes.

One big question this thesis poses is: whether legal-regulatory logic is the appropriate language in which to model the microeconomic relationships that largely determine economic outcomes? Notwithstanding the numerical-mathematical nature of finance, and finance is an important component of the model set out here, finance is only one factor that is taken into account. Financial theory itself generally builds on some underlying economic relationships that it treats as random, or at least independent. In many ways this study is an attempt to broaden financial logic, though in a way that breaks the constraint on only dealing with quantitative data. This is only a tentative

first step, so the legal-regulatory logic is relatively unsophisticated, and will almost certainly benefit from more detailed subsequent work. It may also turn out that there is a more important role for mathematical models than the question of the incorporation of legal-regulatory logic into economic study may at first suggest. Nevertheless, no attempt has deliberately been made to express the ideas contained here in a mathematical form, at least in part to suggest that it is the logic that is more important than the language, although these are not completely separate things.

The relative extent and impact of the failures of Enron and Argentina, where Enron was dealt with in a relatively orderly way, with less serious broader consequences, suggests that the environment in which firms operate in the USA benefits from much greater systematic support compared to economic relations involving sovereign states. Indeed both official bodies, such as the IMF, and influential advocacy groups, such as Jubilee Research, are proposing major changes to the legal environment for sovereign financing (Krueger 2002; Pettifor 2002). So although it is an important subject of much current practical relevance, because of the currency of the subject and consequent difficulties in dealing with the limited evidence readily available, the environment in which Argentina contracted for its debt will not be investigated. Indeed, the point is that it is not this environment that should be of primary concern in understanding the economic outcome, but the institutional structures within the country itself. The focus of this study is firmly in the business arena, focusing on the legal and regulatory environment in which firms of all sizes operate, in relatively well-developed economies for their time. Subsequent studies could make explicit comparisons with relatively less well-developed economic environments.

Hopefully these general remarks help the reader by setting the context in which the subsequent argument is set out. Now is the point to move on from motivational issues set out in a discursive way, to a more specific, and positive, mode of argument.

1.2.2 Legal and Financial Logic

The following five elements combine to form a coherent logical framework, at the micro level, which shows how innovators are able to bring to market new products and services, despite the necessity for relatively large initial investments. The framework is set out starting with the profit opportunity, and shows how entrepreneurs are able to draw on resources across the economy to tackle them, given the availability of specific legal and financial mechanisms. This suggests that once these mechanisms exist the whole economy is able to experience growth.

Innovation

The essence of business activity is the search for profit opportunities. In a competitive market no business can afford to stand still, or another business will find a way of better serving its customers, and win them over. In addition, this dynamic activity is forever opening up opportunities for goods not previously produced. These improvements, or at least hoped improvements, whether in products or services, amount to innovation.

Innovation is on the whole a positive benefit because, broadly construed, it is the only way to develop the resources that help counter the detrimental side of the fluctuations that inevitably accompany human life. Accumulated successful innovation is characterised as economic growth; but that is getting a little ahead of ourselves in the story, more on this later.

Despite its apparent benefits, innovation is however expensive, disruptive and uncertain in its results. The extent to which it is able to generate a profit in the future is subject to a whole myriad of factors, and practically impossible to predict.

Investment

Innovation, in the broadest sense, including all the initial research, development and marketing, can absorb a lot of people's time and resources, with no immediate return. In other words, realising an innovation requires

forgone consumption, or saving, but this can be justified on the basis of an expectation of greater future consumption.

In most cases the innovator themselves, either an individual or a firm, does not have enough readily available consumption to be forgone. Time normally devoted to production will need to be devoted to the innovation, in addition to resources produced by others, to develop the service or manufacture the product, and introduce it to the market. So other people need to forgo immediate consumption, so that the resources they produce can be invested in the innovation.

Forgoing consumption will in many cases be seen as a positive benefit; if in doing so people are building up a claim on resources they are confident they will be able to call on in an uncertain future. The opportunity for greater consumption in the future than currently forgone will further improve the attractiveness of the opportunity. The key word here is 'confident'. The next two elements, governance and finance, provide the mechanisms necessary to provide confidence, and lie at the heart of what could be characterised as capitalist society.

Before proceeding, one additional point to make is that the innovator will, or at least should, know their own market better than any investors, but is not in a position to compare opportunities in their market with other markets, because that information is not readily available. Their expertise, in the form of market specific information and further information that will become available in the course of development, is central to the success of the venture, and will require the continuing commitment of the innovator to bring the venture to a successful conclusion. In contrast, the investor need have no specific information, other than confidence in their future claim on resources, and could even exchange this, that is sell it, if someone else was happy to do this for a mutually agreed amount of immediate consumption opportunities. It is this information and commitment asymmetry that

constitutes the property rights approach, which is set out in (Hart 1995). However, in the model being described here, incomplete contracts are not just a reasonable or empirically justified assumption, as Hart suggests, but necessary for the innovator to take the best course of action as new information becomes available. Furthermore, the systemic elements of governance and finance set out next have not until now been linked with the property rights approach.

Governance

So an investor sees an opportunity to forgo consumption in the short term, in return for consumption in the future. But how do they know, given an uncertain future, that their investment will come good? Two elements of social structure are needed to provide the confidence necessary to make this investment, namely institutions facilitating commitment and credit. The innovator needs to be able to make a credible commitment, in the terms of (Schelling 1980, 24), to give up assets or future revenue in return for the investors forgoing consumption, even if the innovation fails to produce gains. This puts the responsibility for judging the primary risks in the hands of the innovator, who has the best information. Of course it also gives them all the upside of the opportunity, after the agreed amount of return to the investor is taken into account. The assets or future revenue that the innovator commits is their capital. Because modern finance theory recognises that the value of an asset is its discounted future income, this amounts to the same thing. In practice, the asset will need to have a market value, or future income be secure, and it is the function of banks to assess this.

In a national economy this governance mechanism is provided through a legal regime that is able, in the case of dispute, to effectively determine and enforce a previously agreed contract. This also ensures that the innovator will have to return to the investor no more than they have committed to.

But how does the investor know what level of return they should expect, and how this relates to the risks they are bearing? This requires an efficient capital market, as the next section explains.

Arbitrage-free Finance

As previously noted, an investor cannot really judge the likelihood of success of a project. What will give them confidence is a capital market that minimises information asymmetries, so that in effect no investment is better than any other, so long as the innovator has the capital to secure the investment. The responsibility again is with the innovator, in this case as to whether they believe they will make a sufficient return given the rate of return the capital market requires.

It is to fulfil the requirement for shifting consumption opportunities from investors to innovators that this framework justifies the widely accepted, but not well justified, efficient-market hypothesis.

A brief caveat should probably be mentioned. Efficient markets do not happen 'by accident,' but depend on traders and market makers being able to make gains from arbitrage opportunities at the margins. Nevertheless, the basic logic still holds.

Growth

We now broaden our perspective, one step wider, to incorporate all innovation possibilities, as well as the capital market; in other words, the whole economy. Economic growth emerges as the aggregate of new consumption opportunities exercised less the opportunities forgone as savings. General equilibrium and intertemporal preferences work to match overall consumption available to be forgone with innovators wishing to commit to provide investment opportunities.

The connection between investments and returns is not straightforward however. One big factor is the ability of the economy to absorb new products

and services, particularly as in most cases this will be in place of existing ones. There are two elements to this. One is the extent to which vested interests have the power to intimidate innovators, or unreasonably obstruct them. One way to mitigate this is through the participation of those with social and political power in the capital markets. This affects the innovative push. The corresponding effect is consumer pull, in the sense of the willingness of people to try consuming new products and services, and continue consuming those that provide greater benefits at a given price. Both of these effects are difficult to quantify. Nevertheless, they are probably fairly stable over the course of a few years.

The system becomes self-sustaining as successful innovators and investors continue to forgo immediate consumption opportunities for a claim on future consumption opportunities. Furthermore, stable growth occurs because the focus for change is on products and services with greater utility, not simply a shift of market power from one economic group to another. In addition, the attraction of forgoing consumption becomes not simply to give a degree of reassurance for the uncertain future faced by every individual, but the reasonable expectation that future consumption will be significantly greater than the current consumption forgone.

1.2.3 Economic Consequences

The previous sub-section set out to explain clearly and concisely the connection between the different elements that form a self-supporting logical model of a whole dynamic growing economy. This sub-section shows how this framework accounts for four commonly observed economic phenomena, namely output growth, size of economic units, financial sophistication and cyclical fluctuations. While significant in their own right, together they also help provide a more sophisticated description of economic growth than the simplistic measure of income growth used in contemporary macro models.

Output Growth

While economic growth has a number of aspects, the growth of goods and services available for consumption is the central one, and plays a central role in this model. There would clearly be some connection between people's willingness to trade and the existence of market-supporting institutions, that would fit in with conventional microeconomics, and the partial-equilibrium approach in particular, but unlike here it is far from explicit.

Size of Economic Units

Another issue that partial-equilibrium microeconomics is widely noted as failing to address is the existence of the firm. Logically, it even suggests that the most efficient, and therefore optimal, structure would be for individuals to simply purchase goods and services as they need them, at the market determined price.

Coase's explanation for the existence of the firm, as being due to transaction costs, carries some weight, and was one of the sources on which the property rights approach initially drew (Coase 1937). Nevertheless, this approach largely fails to recognise the dynamic nature of the activities pursued by firms.

The approach set out here brings the firm's continuing search for new profit opportunities to centre stage, and provides a far more compelling explanation for the existence of the firm. In the approach presented here the firm is to a large extent a commitment device, in two ways. First, it binds more than one innovator, each of whom brings complementary skills and resources to bear on the opportunity offered. Secondly, it demonstrates the mutual commitment to the enterprise as a whole of investors who are unable or unwilling to provide the whole of the initial resource requirement themselves.

This evidence all fits with the explanation that a market cannot operate if an initial investment is required and success depends on subsequent commitment, because it is impossible to price individual elements. This also suggests that as larger investments are needed to achieve more costly

innovations, albeit with correspondingly large benefits, and credit markets develop that are able to facilitate these investments, then larger economic units will form.

Financial Structure

Although many smaller firms still have relatively simple capital structures, for most large firms this is no longer the case. Complications to the framework occur when share ownership has limited liability, and particularly when equity becomes tradable; because the owners are no longer owners in the sense that they are committed to the return of loaned capital. Rather, they are committed to the extent of the capital they have invested in the firm's equity, in the expectation that they will share in its future profit. Equity becomes a proxy for expected future income, and forms the capital on which debt is secured. This clearly gives equity holders an interest in the management of the firm, but their limited commitment is also reflected in an increased role for other stakeholders, as is increasingly being recognised.

However, these growing complications do have the benefit that they enable a closer match between risk and return, and this benefit will outweigh the cost of the additional complications involved as larger investments are undertaken. So another aspect of growth is an increase in the variety of financial forms.

Cyclical Fluctuations

Despite the difficulty quantifying the socio-cultural factors that affect the degree of innovative push and consumer pull across the economy, as noted above, these are likely to be relatively stable factors. More noticeable medium-term variations in the capital markets will be due to changes in intertemporal preferences. These will be quickly communicated because of the efficient nature of the capital markets, and will be to some extent self-reinforcing. This corresponds with the commonly recognised role of sentiment in capital market pricing. Fluctuations in financial markets will in turn be communicated to the real economy through shifts in investment patterns.

This suggests that as credit markets develop there will be a tendency for output growth to cycle, simply due to these endogenous factors, rather than in response to non-market factors such as wars and epidemics.

In many ways, the framework set out here asks more questions than it provides answers. What it is trying to do, however, is to provide a better, more comprehensive and more coherent, way of looking at a wide range of questions.

The commitment-credit model sketched out here, while not set out in a formal mathematical language, creates a bridge between institutional and financial economics. This provides a framework for addressing many of the issues faced by managers and policymakers, to the extent that it may make the conventional partial-equilibrium approach to microeconomics as redundant to economic education as it is to economic practice.

1.2.4 Scientific Tests

The model just set out therefore claims to provide a better explanation of microeconomic phenomena than the neoclassical model. The challenge in this thesis is to substantiate that claim. This sub-section sets out how this will be done. The central issue is in what way this should be done, and the approach here is to take a scientific perspective. So the first question is: what does this mean? Once this question has been answered the structure of the thesis will fall into place.

The model posits that a credible contracting mechanism is necessary for channelling potential savings through a functioning financial market and that the three are necessary for channelling innovative capability into growth. This implies that a change in contracting mechanisms will precede a change in financial markets that will, in turn, given that human beings have an inherent innovative capacity, precede a change in growth. And growth breaks down into four specific factors, namely the rate of income growth, an increase in the size of the unit of economic production, an increase in the variety of

investment vehicles and that the rate of income growth will display cyclical variation. The question is how can this model be tested?

Given the fashion in social sciences at the moment, the first thing to spring to mind is to look for a statistical test. But this is problematic, for two reasons. First of all it would be necessary to put together a big enough sample to give significant results. The second issue is that while a credible contracting environment should be observable, it is not readily apparent how it would be quantifiable. This however, is not as problematic as it may first appear. Popper's analysis, considered in more detail when assessing the contribution of this work in the final chapter, shows that what is of primary importance in assessing scientific theories is the existence of falsifiable, empirically observable, basic statements. It is impossible to ever go from the specific to the general, so even if the validity of basic statements could be directly established in every observed case, never mind in a statistically significant sample of them, these cases would never 'prove' the theory. So this study focuses on directly testing a small number of hypotheses through direct observation, leaving wider statistical tests to later studies.

Given the causal mechanism the model describes, an obvious test is to find a change in credible contracting mechanisms and determine the extent to which that was followed by a corresponding change in growth. But it is far from obvious where to start looking for changes in credible contracting mechanisms, so this is a problematic approach. The opposite approach is to examine the extent to which changes in growth were preceded by changes in credible contracting environments, and that is what this project did. It would be possible to present situations found in this way as being selected on the basis of changes in their contracting environments, and then looking for growth impacts, but this would be somewhat disingenuous.

Two specific shifts in the rate of economic growth spring to mind, namely the industrial revolution in England and the growth of large corporations at the

end of the nineteenth century in the United States, sometimes labelled the managerial revolution. Both of these are very problematic from the neoclassical perspective, because it suggests that economic activity is more efficient and therefore better socially when it is undertaken by as small an individual unit as possible, while both these cases at least on the face of it saw the opposite trend, with new larger organisational forms competing more effectively than older smaller forms. The first case saw the rise of the factory system and the second saw the rise of corporations employing tens of thousands of employees and capitalised at hundreds of millions of dollars. Partly because of the problematic nature of the neoclassical model, both these shifts in the nature of economic activity continue to excite controversy as to the underlying nature of the change and its causes. So, if the model proposed here was able to provide an explanation of both of these shifts, it would also make a considerable contribution to these debates.

While these two large-scale historical developments define the consequences for which we hypothesise that the causes were developments in institutional commitment and credit mechanisms, it may not be obvious where we should look to observe these previous changes. It is perhaps more obvious in the first case, where the focus is a whole national economy. In this case the legal environment for settling contract disputes, and then the overall financial system, are the areas to focus on. In the latter case it is the managerial performance systems, followed by the private capital allocation mechanism, that are the areas of concern. In the latter case accounting techniques provide the framework for setting and assessing managerial performance agreements, and will be the focus for investigation.

In large part because of the continuing controversy around these historical shifts, there is a considerable body of historical scholarship providing a substantial number of easily accessible secondary sources, based on thorough examinations of a wide variety of primary sources. These secondary sources provide an excellent supply of empirical data on which to test the relationship

between commitment environments, credit markets and growth. In particular, because a record of the agreement is an important element of an agreement, as it reduces the possibility for future disputes, many of these may still be extant, and should provide a considerable body of historical evidence. In fact, it quickly becomes apparent that there is a more than adequate supply of historical evidence available to use in this project, particularly in the form of legal and accounting history based on extensive primary sources.

Using secondary sources clearly presents issues as to the bias introduced by their authors in the selections they make to present their arguments, but this should not be a significant problem. This is not primarily a historical work. It is focused on social scientific theory, and simply makes use of historical evidence to test the theory it puts forward. Because the historical evidence is used in this way, and there is little direct engagement in the arguments made in the historical works, the evidence drawn from them can to a large extent be considered as a random sample of the extant evidence.

We therefore have hypotheses to test and readily available evidence to use to carry out the test. Our hypotheses are that shifts in the structure of the economy leading to greater economic growth in the industrial and managerial revolutions were preceded by improvements in the mechanisms available for making credible commitments and in the markets for financial credit. And a variety of secondary historical evidence exists that provide descriptions of developments in legal and accounting environments, in financial markets, and economic growth.

A theory can be thought of as consisting of a central logical model and a well understood approach to applying the model. It is on this basis that this study asks to be judged in its contribution to improving the understanding of the microeconomic aspect of the world. Although this project is largely motivated by theoretical concerns, and much of the time was taken seeking to address them, once the model had been formulated it was relatively quick and easy to

present. The bulk of this final report shows how this model can be interpreted in terms of observations. It may be unnecessary, given its limited conceptual elaboration, to point out that the model should not be judged on its technical sophistication, but on its explanatory value. It is in the nature of this explanation that the model claims its value, and the considerable amount of observational evidence presented here gives considerable additional depth to the theoretical concepts. This is as it should be. For any successful science, its central concepts at first appear over-simplistic, and the observations it chooses to make concern only a seriously limited subset of those that appear to the untrained eye as potentially relevant. Where it gains its strength is in the interaction between well-defined concepts and a wide range of actively pursued observations. However, a theory can also be assessed on the importance of the phenomena it explains. In giving a clear explanation of the major changes in economic activity in recent human history the theory further strengthens its claim to make an important contribution.

While this study now has its central focus, there are three additional issues that can be addressed, and will help to establish the scientific credentials of the theory. First, the definitions of growth given here are more than the simple income growth normally taken to be the primary measure of economic development, so the extent to which this richer elaboration of growth is in line with the historical evidence will also be examined. Secondly, a small number of specific examples of particular businesses based on significant technical innovations will be examined, to see to what extent they were able to survive and grow given first the general environment in which they operated and secondly, and possibly more problematic given limited evidence, the extent to which where this was possible they were able to make use of the commitment and credit mechanisms available. Finally, it will also be an interesting test to see if there is evidence for the lack effective mechanisms in eras prior to those in which sustainable economic growth occurred. This will also help establish the extent to which developments of commitment and credit mechanisms were genuinely novel.

1.2.5 Presentation of Results

The structure of this thesis follows naturally from the hypotheses that test the commitment-credit theory. Over half of the text details the historical evidence, which turns out to confirm the theory. The subsequent chapters assess the implications.

Given the historical nature of the evidence presented, it is natural to present this in a temporal narrative. So the next chapter examines the evidence for the lack of effective commitment and credit mechanisms in ancient and medieval Europe. This also provides the background to the subsequent chapter to understand what was different about England in the run up to the industrial revolution.

The first substantial positive test of the theory is carried out in Chapter Three. The development of the legal environment in England, particularly as it relates to contract law, and in the domestic credit market, up until the eighteenth century, is set out in some detail. Following this the evidence for income growth, the growth of firms, increasing variety of financial instruments and the start of a business cycle in the eighteenth century is examined.

Despite becoming independent in 1776, the United States inherited much of its financial and legal systems from England. The end of the nineteenth century however saw an unprecedented restructuring of industrial firms. Although perhaps not so much studied as a particular historical epoch as what is known as the industrial revolution, the managerial revolution is to some extent easier to pin down. Between 1895 and 1904 over 1,800 firms disappeared into ninety-three corporations. The fourth chapter examines the developments in managerial accountability and corporate finance that preceded this, and then examines the issues of income growth, growth in internal organisational sophistication, internal capital allocation and output cycles for the large corporations.

The following chapter moves towards assessing the implications of this work. Neoclassical problems have not gone unnoticed and modifications and alternatives incorporating a central role for governance in some form, of a largely un-mathematical nature, have been proposed. Although fairly widely known, they do not seem to have led to changes in the way the neoclassical framework is presented. Nevertheless, partly to help assess the contribution of the work presented here, four significant alternatives to the standard neoclassical model will be discussed in the light of the outcome of the tests. These are, first the property rights interpretation of Western European history of North and Thomas. Next the transaction cost theory of Coase, followed by a further elaboration of this by Williamson. Fourthly, the later institutional theory of North is set out.

The issues dealt with in the commitment-credit model at the micro level, at first sight, mirror those dealt with by macro models, and in some ways this theory could offer similar advances over the neoclassical model, at the micro level, as the Keynesian model did over the neoclassical model at the macro level. So some of these macro issues will also be introduced, to help understand the implications of the contribution made by this model.

The implications of the commitment-credit theory are the subject of the final chapter. The first part of the chapter considers the conceptual nature of scientific enquiry, and the implications of this project for microeconomic research in particular. The second part of the final chapter discusses the social implications, first within the economics profession, then some broader practical implications. Some of these may feed through additional macroeconomic work to public policy, but there are some more immediate implications for microeconomics and management studies, in particular in terms of what should form the conceptual focus for studying the management of large organisations.

The framework provided here provides a much clearer link between management studies and microeconomics than currently exists. Links to other related subjects such as law and both accounting and finance also become apparent, in a far more transparent way than has been possible to construct with the neoclassical framework. What may seem to disappear is the framework as a foundation for macroeconomics, but the conventional focus of macroeconomics over the past few decades has itself come under considerable attack in recent years. Policy concerns have shifted away from attempts to manipulate aggregates such as money supply, employment and public spending, towards such 'micro' issues as competition regulation, banking supervision and public sector governance. The framework provided here provides a much more natural conceptual structure for dealing with these issues. So, as well as exploring possibilities for mathematical formalization, the natural follow up to this project appears to be a macroeconomic theory of governance that is able to bridge the gap.

Chapter Two

HISTORICAL BACKGROUND

It is apparent, from even a cursory knowledge of history, that there was only ever very modest sustained economic growth prior to the eighteenth century. But it is not immediately clear why this was the case. So providing an explanation of this can be construed as a test of the commitment-credit theory. This chapter examines two pre eighteenth-century historical epochs, to see to what extent commitment and credit mechanisms existed, and how this affected economic growth; namely ancient and medieval Europe. Of course, while providing a test of the theory, the details presented here also describe the historical background to the first appearance of sustainable growth, and will help to emphasise the distinctiveness of the social phenomena that enabled it, as well as showing that the governance mechanisms to enable commitment and credit did not just suddenly emerge into being.

The twin purposes of theory testing and subsequent comparison therefore provide the selecting and organising principles for this chapter. Authoritative secondary sources provide the evidence. Nevertheless, some caveats do apply. First of all, for the ancient world in particular, primary sources are far from abundant. However, the focus is on areas relatively well attested by sources. While not conclusive, an evolutionary argument would suggest that those regimes that left the most substantial evidence were the most successful of their contemporaries. Perhaps the most important caveat is that there is no reason to expect that the way people are brought up to categorise their world now is not substantially different from in the past, which makes understanding ancient societies that much more challenging. Indeed, this thesis contends that it is changes in the nature of relationships between

individuals that is at the heart of the emergence of growth-enabling governance.

2.1 Ancient Greece and Rome

Evidence of commitment and credit mechanisms require two things of a society, first a sense of the individual as having a degree of autonomy, and second a physical record. Without the first the concepts make no sense; without the second there is no evidence. On both counts we begin with the Greeks. To a greater or lesser extent all known human societies, almost by definition, could be fruitfully examined, but the Greeks are the first civilization to provide a wide range of written evidence that both directly, through actual documents involved, and indirectly, through their literary and historical sources, demonstrates a multifaceted but relatively consistent conception of the individual in their society, and the existence of mechanisms used to invoke what we call contractual commitment and commercial credit.

2.1.1 Commitment

Most of the elements for facilitating the creation of contractual commitment devices did emerge in some form in the ancient world. The Greek contribution was fundamental, and within their world Athens for our purposes was the most advanced, because they invented the concept of a democratic society. That is, a society where its members set the standards to which everyone must adhere, and the sanctions to be applied to those who do not. Despite the limited nature of Athenian democracy, which excluded women and slaves, it is in stark contrast to the hierarchical structure of most human societies, where those higher up the tree make the decisions for those below. The Romans then advanced on the Greeks by abstracting this principle, and beginning the development of social systems governed by the rule of law. Necessity was largely the mother of this invention, because where Greek city-states never moved beyond a size where communal decisions

could be taken by a significant proportion of the community, Rome, even at the end of the republic, ruled over a substantial empire.

Greek society did develop the concept of contract, but contract law never developed into a recognisably modern form of legal practice, largely because in practice the resolution of disputes depended largely on the reputations of the individuals involved. The Athenians tried to stand firmly by the principle that parties to a contract should explicitly agree and state in the contract the provisions to which they wished to be bound. Although in many ways a laudable aim, as later more sophisticated societies found, it presented a barrier to developing an effective system of contract law. This depends on making the contracting process manageable, by developing widely understood principles for dealing with the inherent complexity and unpredictability of human life, which are taken to hold, unless explicitly ruled out in any particular contract. Further factors of Athenian legal practice acted as additional constraints. One individual could only take legal action against another. Not only did this present a problem in forming any sort of company, it also made partnerships difficult because action could only be taken by or against one partner at a time. It has been suggested that one effect of this was that in a cooperative venture there would only be one named individual, who would be the least wealthy, even though their actual involvement in the venture may have been relatively small, if indeed they were involved at all other than as a proxy for contractual purposes (Cohen 1992, 170).

By the fourth century BCE contracts were written, but up until this point courts demanded oral testimony, and written material was of little value, so contracts were only made verbally (Cohen 1992, 178). This obviously limits the evidence we do have of specific contracts, although some direct evidence does exist from records of court activities, particularly speeches by well-known orators, that there were relatively sophisticated financing contracts for maritime trade.

All things considered, while effective practices could be developed by individuals, and diffused through personal instruction or copying others' observed behaviour, there were no institutional mechanisms for accelerating this learning. Even if one accepts the most positive interpretation of the evidence we have available regarding its sophistication, institutional limitations severely constrained effective development.

The Romans, although they acknowledged their debt to the Greeks, developed the basis of a system that up to this day forms the foundation of one of two widespread legal traditions, one which remains particularly strong in continental western Europe. The foundation of their legal system and subsequent developments was the Twelve Tablets, produced by a commission set up in response to popular calls for a codification of the traditional laws, similar to that of Solon produced in Athens in the early sixth century. A popular assembly in 450 BCE approved them. These dealt with a wide range of issues, focusing on potential areas of contention. In particular, they dealt with details of legal procedure (Stein 1999, 4).

Roman law gained much of its effectiveness from its procedure, which had two elements. The first was the application of legal principles by an elected magistrate, the second the establishment of facts by an independent jurymen. A litigant first had to appear before the magistrate with his opponent. The opponent would presumably go along with this, as there were relatively straightforward ways to compel him if he was reluctant. The magistrate would decide whether there was a case to answer in legal terms, and if there was, what the factual question was that would determine the outcome. A mutually agreed juror was then selected, if necessary using a procedure that was demonstrably fair, where each party eliminated one person from a list of eligible citizens until only one remained. The magistrate would then make a proclamation appointing the juror with his terms of reference, setting out the appropriate legal 'formula'. The juror would then take whatever steps he felt were necessary to determine the relevant facts, and then deliver his verdict.

When this procedure was first instituted, the magistrates were the two consuls responsible for every aspect of governing the city. However, in 367 BCE a new office was established, that of praetor, and someone was elected to the post annually to deal exclusively with the administration of justice (Stein 1999, 8). It was not expected that the individual elected would have any particular expertise in the law. The tradition developed that at the beginning of the year of office a praetor would publish an edict setting out the formulae he would grant, and it was then for individuals to apply for a formula in regard to a particular situation. In theory the praetor never made the law, but simply set out what the law was. However, the edict could and did change from year to year, so the effect was that the law developed in response to the types of cases that arose.

As Rome's sphere of influence expanded and the formal legal framework grew in sophistication, the system developed. A second praetor was elected annually from 242 BCE, and soon afterwards a class of legal experts, jurists, emerged. Neither the magistrates, nor the jurymen, nor for that matter the litigants or defendants, had any particular expertise. Although not having any formal role, jurists were available to advise any of the parties, and at first they were not paid because they regarded their work as a form of public service. They effectively became the custodians of the law (Stein 1999, 13).

This situation was largely maintained in the early years of the Empire. Augustus in particular was anxious to preserve the façade of a republican constitution. But not surprisingly the praetorian edict had reached the point where it hardly changed from one year to the next. And in the early years of the second century the Emperor Hadrian ordered it put into a permanent form. The two centuries leading up to this are regarded as the high point of Roman law, known as the classical period. During this period expert jurists debated problematic cases that arose, either in reality or hypothetically, and these debates shaped a remarkably rational and consistent legal framework.

Alongside the maturing of the central principles was a move towards systematisation. At least at first, this was as an aid to legal education rather than legal practice. But by the beginning of the second century two leading jurists, Paul and Ulpian, wrote great commentaries, synthesising the work of their predecessors. Ulpian made for the first time a clear distinction between public and private law. Public law addressed matters to do with the state, and private law addressed relations between private individuals (Stein 1999, 21).

It is perhaps not surprising that as the system became increasingly complex it also became increasingly bureaucratic. One significant change was the introduction of professional judges, often trained jurists, who took on the roles of both magistrate and juryman. This was a gradual occurrence, with the two types of court existing through most of the Empire until the formulary procedure ceased to be used, officially in the fourth century, but probably in practice the previous century (Johnston 1999, 122). Alongside this, the orality that had remained an important element of the process gave way to exchanges conducted in writing. One generally positive innovation was the introduction of an appeals process. All this placed far heavier demands on the time of legal experts, and the pendulum of development swung away from elaboration towards systematisation and simplification (Stein 1999, 24).

Roman contracts took two forms, formal and informal (Johnston 1999, 77). The formal one, *stipulato*, was made orally. The promisee would ask a question, and the promisor would reply with an exactly corresponding answer. Its terms were interpreted strictly and everything needed to be stated explicitly for it to be a part of the contract. Informal contracts covered a variety of specific situations, and each type of situation required its own type of contract. The emergence of the consensual contract of sale, *emptio venditio*, by the end of the second century BCE, overcame some of the disadvantages of the *stipulato*. It required a defined object and price, which clearly covers many situations, but excludes those where the price may be subject to change for any reason, by reference to a future event for example, or because the

object is not specifically identified, as with a generic product such as corn before the specific measure has been metered out. Given the required elements, several legal consequences followed automatically, such as the seller warranting title to the goods and the quality of the goods, and once agreed the risk of accidental loss or damage passed to the buyer. In more general terms, the concept of good faith developed, and was applied, by the early second century CE. The effect of this was that the law automatically kept pace with customs of trade and commerce (Johnston 1999, 81).

Given the important role of banking in the economy, described in the next subsection, the law evolved to deal with significant issues that would arise from time to time. By late classical law it had become accepted that, as a consequence of good faith, if a banker made use of money placed on deposit, he was obliged to pay interest (Johnston 1999, 87). Security against borrowing generated a considerable juristic output, suggesting it was a very real issue. Much of this related to personal guarantees, suggesting that this was creditors' preferred source of security, but real security was also significant, although the scope for confusion over prior charges means that it was less useful and versatile than it might have been (Johnston 1999, 94).

Moving from contracts related to specific actions, what of the legal basis for principal-agent relationships? As a precursor to this it is necessary to set out, briefly, significant differences between Roman society and our own regarding interpersonal relationships. The primary difference was that in Rome the family rather than the individual was seen as the basic unit of society. In legal terms this meant that the senior male ascendant, the *paterfamilias*, owned all the family property, and everything acquired by his dependants automatically vested in him. As well as descendants the family also included slaves. One possibly surprising difference was that marriage did not alter this family relationship, so at least in property terms divorce was fairly straightforward. Not surprisingly perhaps, there was a mechanism that made this property situation a little less restrictive than it otherwise would have been. This was

the *peculium*, where some amount of property was made available to a dependent descendent or slave. This remained the property of the paterfamilias and could be withdrawn at any time, but until that happened the person to whom it was granted had full powers of disposal over it (Johnston 1999, 31). There were legal avenues available for releasing descendants or slaves from the control of the paterfamilias, but certainly in the case of children this does not appear to have been the regular occurrence. It was largely on the elaboration of the *peculium* concept that principal-agent relationships developed.

What follows is based less on evidence of actual practice than on the extent to which the issues are addressed by the jurists, because of the lack of direct evidence of the former, while the latter does not seem unreasonably disconnected from practical application. The effect of the *peculium* was that the paterfamilias would still be called as the defendant in any legal proceeding, but would be bound by the actions of a dependent, to the limit of their *peculium*. This actually had a benefit of course, that the limit of the action was the extent of the *peculium*, so introducing a form of limited liability. There were cases where a principal could be sued for the actions of an agent, one of these related specifically to owners being responsible for the actions of their ship captains. In these cases there was no limit to liability. For a dependant in receipt of a *peculium*, the extent of the liability rested on the knowledge of the paterfamilias. Clearly, there could be benefits from not having direct knowledge, and one way of helping establish this was by having a dependent with a *peculium* appointing the agent, a possibility explicitly acknowledged by the jurists (Johnston 1999, 104). Employing agents who were not dependents did bring additional complications, although these were reduced during the second century CE. There is evidence of partnerships, *societas*, although it is more evident as an arrangement under which co-heirs would continue to operate an inherited business as a single entity, than for more specific business opportunities, although these were not ruled out. One document sets out a partnership arrangement, but in the form of a *stipulatio* (Johnston 1999, 107).

Part of the reason for this may be that the partnership was only relevant for the partners. Any action by another party had to be in relation to one specific partner.

One other relevant issue addressed by Roman law was bankruptcy. A creditor could take action in two ways, against a person, or against their property. Action against a person appears to have become less common. A creditor could imprison a person until he had worked off his debt, but although there are provisions in the Twelve Tables for this, its application is unclear. More clarity prevails for action against property. Detailed consideration by the jurists and continuing development through the Empire suggest that this was a regular occurrence, for example elaborations were introduced to treat all creditors on an equal footing (Johnston 1999, 109). With no concept of separating the individual from their business, the individual was wholly liable, as well as being subject to subsequent civil disabilities, and social stigma. As this was in general the only remedy open to an unpaid creditor, it could create difficulties for both sides, but perhaps in practice this would have encouraged mutual agreement without court involvement.

From this brief summary it is apparent that Roman law had at least begun to address the central issues of contracts as commitment devices for commercial activity. The next section describes the further elaborations developed during the middle ages, but Roman law continued to be the foundation on which subsequent developments were built. The next question to be addressed is to what extent did this foundation allow financial markets to develop.

2.1.2 Credit

To some extent our story ends before it really gets started, because neither the Greeks nor the Romans, according to one of their most generous analysts (Andreu 1999, 147), distinguished between loans for investment and loans for consumption. Much debate has gone on as to the extent to which ancient economies functioned in similar ways to modern ones. While Finley, on

whose authority we will also draw, has made the case for a significant difference, Cohen and Andreu tend to the other point of view. However, in all cases, what will be drawn on here is the specific nature of the mechanisms the evidence shows were available, rather than any assessment as to their modernity.

The lack of any evidence directly or indirectly of a distinction between loans for investment and loans for consumption suggests that the role of credit markets in allowing innovators to create wealth was not appreciated either explicitly or implicitly. So they failed to function effectively as a governance mechanism harnessing innovation to growth, and in itself provides an explanation of the ancient world's failure to achieve sustainable significant growth. Nevertheless, well-documented credit mechanisms did exist, and these will be set out. Direct connections between Greek, Roman and subsequent European practices will become apparent, and the issues dealt with by the ancients do shed light on the nature of effective governance mechanisms, particularly in regard to problems that can arise.

Money, the precursor to any effective financial credit mechanism, was a Greek invention. The use of standardised pieces of precious metals goes back much further, but coins were invented around 640 BCE in Lydia, in what is now Turkey, but was a city state in the Greek tradition (Davies 1994, 63). Archaeological evidence clearly shows developments leading to the coin around Lydia, where other areas of the Greek world only reveal fully developed coins. Nevertheless, the take up of the idea was very rapid. Around half a century later they were widely in use across the Greek world. Although no doubt initially understood as nothing more than a way of identifying a recognisably unadulterated standard amount of precious metal, the concept of a unit of currency as a measure of value in itself, rather than simply another commodity, albeit a portable, easily transported and traded one, opened a new world of abstraction.

There is some evidence that Athenian banking became relatively sophisticated, and dealt with a variety of clients including both wealthy aristocrats and businesses, particularly the maritime traders who among other things kept the Athenian population fed; although the extent of their activities is disputed (Cohen 1992, 8). At least in part, the dispute is a result of the somewhat confusing evidence. So even if credit mechanisms were relatively sophisticated, it has been suggested that the legal framework available for mediating disputes, just described, and the mechanisms for collecting taxes described next, at the least provided incentives for limiting the amount of information put into accessible forms. Clearly, information that was not easily available to contemporaries is even less available to us from our distance in time.

Athenian taxes were effectively levied on visible wealth, and worked on the basis of specific needs for public expenditure being made the responsibility of a particular citizen. Common examples were a ship for the navy, or funding for a communal festival. A process existed for the community to select specific individuals. But if the individual selected could prove in the law court that another individual was wealthier and had not performed as many public duties, the burden shifted to them. While this system makes a lot of sense, given that there was no government machinery, there was clearly an incentive to conceal the extent of one's wealth, which one only had to do from other citizens, rather than the state. This not only provided an incentive to keep bank activity confidential, but because confidentiality was so important to a banker's success, it is possible that this increased the demand for bank-mediated investments. This may have become particularly acute in the fourth century, when Athens had lost her empire, its silver mines were less productive, and while it still remained independent, the tax burden became rather substantial (Cohen 1992, 194). But the source of funds, particularly given the incentives mentioned, is open to a variety of interpretations, so it is not clear how credit arrangements fitted into the wider economy.

In contrast to the Athenian picture, there is considerable evidence that Roman banking did become relatively well developed. The industry existed in two forms, although, as with most social classifications, these are not completely clear-cut. Many aristocrats both loaned and borrowed money from each other, for a variety of purposes. In addition, there was a profession of banker. This was at the level of what would now be called retail banking, formed from those who plied their trade in shops as any other trader. The trade was regulated to some extent, and the law recognised a number of distinct types of transaction, as well as standard practices to which those functioning as bankers were expected to follow. Needless to say, perhaps, these regulations did not apply to aristocratic lending.

Before saying a little more about these two categories, a couple of general points are worth noting. First, while historical evidence suggests that these two types of lending were very common, there were other forms, less common, but still relatively well established. One was lending for sea-borne trade, which also acted as a form of insurance, as interest rates were high, but the lender bore more risks. Another was the large-scale merchant who would often use credit in the course of their arrangements, and would sometimes be found with this as their primary activity. It is also likely that there would have been those who lent to the poor at extortionate rates, though partly because this was technically illegal, and partly because of the illiteracy of the people who participated, not surprisingly, there are no extant written records, so nothing is known about it. The second general point is that lending meant lending coin. There is no evidence that there was ever any other form of money, and this of course placed enormous strains on the financial system from time to time. Nevertheless, the fungibility of money was explicitly recognised in Roman law, so that unlike other cases where subsequent disputes could arise, in the case of coin, title always passed with possession (Johnston 1999, 55).

In Rome, like most societies, heredity played a crucial role in the social structure. A few wealthy families passed control over large areas of land from one generation to the next. On the whole these individuals devoted little time or effort to agriculture, but used the income from it to carry out what they regarded as their social and political duties. Making a political career was expensive. One had to develop supporters among particular influential individuals and the population as a whole, and for this significant expenditure was necessary (Finley 1985, 153-154). Often this meant borrowing money from other families. The return came later, particularly under the Empire for those who became provincial governors, when one was rewarded by a significant proportion of the tax revenue (Finley 1985, 53). Individual civic donations made up a good deal of public expenditure undertaken, and in particular it was the primary source of non-military capital expenditure, because there was no concept of public debt. Of course, the individuals involved creamed off a significant proportion of tax revenue for their own consumption, and lived in considerable luxury.

Roman bankers carried out a number of commonly recognised functions, including assaying coins, foreign exchange, advancing credit at auctions, receipt of deposits and subsequent payments, and the advancing of loans. In many cases they were awarded a monopoly, and had a tax levied. For example, in some cities changing money from large to small denominations was restricted to the recognised monopolies, which charged about five percent in commission, some of which went in tax. A banker was expected to keep a register of accounts, which was required by law to be produced whenever a client was involved a lawsuit. It was not considered to provide absolute proof, but was deemed to be reliable evidence. This at least suggests that bank accounts were relatively common, if not *de rigour*. Interest rates were a subject of public concern, although the evidence seems a little confused as to what was thought reasonable (Andreu 1999, 91). The maximum rate fixed by the state, generally agreed to be around twelve percent, was not the basic rate, and there is considerable evidence of loans below it, often around five to

six percent. There is also evidence, not surprisingly less common, that rates over the theoretical maximum were charged from time to time. There was also an expectation that this would be simple interest, compound interest was sometimes allowed, but frequently forbidden (Andreu 1999, 92), betraying the underlying lack of appreciation of financial logic.

In regard to payment mechanisms, there is some evidence of the use of cheques, but they are always direct, so never becoming negotiable and constituting an extension of the concept of money (Andreu 1999, 42). As one would expect, from time to time there were shortages of coin, and particularly when it was in demand for special taxes for example, the price of land could fall, that is the value of the coin increase, as demand for coin increased. As always, this inflexibility in the money supply would severely and unnecessarily magnify the problems of individuals who were unfortunate enough to be experiencing temporary cash-flow problems, as well as accelerating the consequences for those in fundamentally unsustainable economic positions.

So, while the invention of money helped the notion of credit to develop, the limitation to precious metals also limited development. Nevertheless, the next question is to assess the extent to which the commitment and credit mechanisms that were available in the ancient world could have enabled sustainable economic growth.

2.1.3 Consequences

The fate of Athenian and Roman civilization clearly demonstrates that there was no such thing as sustainable economic growth in the ancient world. Athens remained an important city well into Roman times, and Roman civilisation could trace a continuous existence through to the final fall of Constantinople in 1453 CE. In many respects, and certainly in its transfer to the East, it was far from a static civilization. Developments were taking place all the time. Even in the few details sketched out above, a huge creative potential can be seen. And the areas described by no means stand out as

unusual. In architecture and civil engineering the Greek and Roman achievement stands to this day. Roman roads and aqueducts would be significant achievements for modern engineers. Roman and particularly Athenian literature and philosophy is still admired and studied. Their influence pervades western civilization. But they have no direct descendants. They are dead civilizations.

Greek and Roman civilisation grew, reached a peak, or a series of peaks, declined and died. And the fundamental reason for this unsustainability lies in the answer to the question: who benefited from the achievements of the civilization? The answer: the social elite. For the vast majority of people who made up Greek or Roman civilization nothing changed. Agriculture dominated the lives of the vast majority, and despite all the creativity unleashed on what could be called high culture, innovation that would affect, never mind improve, the lives of the vast majority were singularly, and to our eyes startlingly, non-existent. Trade undoubtedly increased, but archaeological evidence shows that trade is a pervasive human characteristic, not limited to common cultural or political areas. But most people would see little benefit. There were more city dwellers, but the vast majority of the inhabitants of Athens and Rome themselves existed at subsistence levels. The water supplied by aqueduct, or bread freely distributed from time to time by rich Romans, was a necessity to life, not in any sense a beneficial luxury.

One benefit under the Roman Empire was a lack of war for those in the empire's interior. But although this could be a benefit, it was probably difficult to recognise at the time. While most pre-modern societies could only support any large-scale military mobilization when imminent hostilities threatened, Rome maintained a substantial standing army. This was a considerable expense, and the government would have appropriated most of any production surplus to basic subsistence. As well as supplying the army, this was used for the massive construction projects and luxurious lifestyles of

a few, that while clearly providing marvels for future generations, meant nothing to the vast majority.

Rome's eventual failure has been ascribed to its failure to ever improve on its productive capacity (Finley 1985, 149). Is it possible to be more specific than this? The lack of governance mechanisms facilitating commitment and credit offers a potential explanation, particularly because, as has been shown, they existed to some extent. To what extent can this be posited as a specific explanation? Does it genuinely improve on explanations in terms of less specific but more pervasive cultural factors?

The unbreakable glass story goes some way to explaining Roman attitudes towards innovation, particularly among the literary elements of society. A number of ancient authors, including the elder Pliny, Petronius and Dio Cassius cite an event in the reign of Tiberius. A man has invented unbreakable glass, and demonstrated it to the Emperor expecting a great reward. After satisfying himself that the man had shared his secret with no one else, his head was promptly ordered removed lest, explains Tiberius, the value of gold be reduced to that of mud (Finley 1985, 147). But what opportunity would the man have had for profiting from his invention? Leaving aside not unimportant issues of intellectual property, could he have secured the necessary capital to develop and market the product, either himself, or in partnership with someone else? The quick answer is yes. He could have approached a banker, and offered whatever property he had available, perhaps his workshop, as security. Failing the adequacy of this he could have entered into a partnership with someone who had sufficient personal or material standing to act as security, on whatever profit sharing arrangement was agreed.

However, while in principle the mechanisms were available, the practical difficulties would have been considerable, if not overwhelming. The scant evidence from actual commercial practice, as opposed to legal possibilities,

suggests two particular difficulties. What little evidence there is suggests that security outside the aristocracy always took the form of moveable goods, such as slaves, precious objects and traded commodities (Andreu 1999, 75). And among the few things known about some loans is that they were relatively short term, for a few months or at most a year (Andreu 1999, 44). This goes to suggest that commercial lending addressed the need for working capital, rather than investment capital. Without any substitute for cash, such as endorsable bills, the seasonable nature of the agricultural products that made up most trade, and the significant times goods were in transit, meant that there would be a considerable need for working capital. The lower risk associated with this type of lending, together with the limited supply of money, would largely preclude the availability of investment capital.

In terms of the partnership contract, further problems would arise. One could speculate that this would have been a particularly unusual arrangement, but then so would the invention of unbreakable glass! More seriously, the difficulty would have been in finding a partner. Relationships were, as we have seen, despite the recognition of some individual autonomy, perceived very hierarchically. So although there could be partnership, this in general is seen in similar roles, rather than complementary ones. A wealthy partner clearly would be unwilling to become the junior partner, but would the inventor want to join the capital supplying partner's establishment, indeed, would it be possible if he did? It is not clear that, between becoming a slave or being adopted as an heir, there was any practical mechanism.

In recognising these practical details we start to move towards the impact of cultural factors. While physical property could be used as collateral for a loan, it was far less problematic, and therefore presumably greatly to be preferred, to use personal standing. But of course one had to have the personal standing, meaning family wealth. To a large extent this is a consequence of the failure to distinguish between lending for consumption against lending for investment. For commercial lending, it would almost inevitably be for working capital.

For lending among the aristocracy, in a sense lending to further a political career was lending for investment, but not in a systematic sense. On the whole, in both cases lending was to overcome a short-term shortage of cash, simply in the expectation that in the fullness of time income would return to its previous level, and the loan could be repaid. Underlying all this was an essentially static view of society. One was born into a certain station in life, and that was where one should remain. Given the lack of sustainable economic growth this was not an unreasonable view to hold, but in this way the failure to grow reflected a failure of expectation as much as anything else.

This essentially fixed view of society would prevent another problem to an innovator, which perhaps reflects Tiberius' concern. To whom would the inventor sell his new products? And if he did sell them, would that not imply that someone else would lose out? Innovation would threaten the whole existing basis of society. And of course that was the big threat to those who felt themselves to be at the top of the pile. So innovation was best actively discouraged. With this pervasive attitude, it would be far from clear where the revenue to repay the capital investment would come from, so difficulties with the culture and the business logic went hand in hand.

In methodological terms, this difficulty in distinguishing, and indeed a genuine interdependency, between specific commitment and credit institutions and more general cultural factors is the continuing problem this study faces. What can perhaps be said, is that cultural and formal practices went together, and both played their part. The formal elements functioned as they did largely because of their cultural background. But, in some ways more importantly, because they can be more clearly perceived, the formal legal-financial mechanisms available had very real consequences, in the extent they either prevented or enabled sustainable economic growth.

The broader point is already established however, that once one knows what one is looking for, in terms of formal commitment and credit mechanisms,

even in distant, scantily attested, civilizations, it is relatively straightforward to identify key elements and the extent to which they performed the function that the commitment-credit model suggests they need to. It is also readily apparent that although the model deals with the concepts in a clear and easily understood way, their actual embodiment in a society is far from straightforward. The ancient world made considerable steps in their development, but still left significant issues to be resolved.

It probably adds little worthwhile to say that the failure of Roman civilisation to sustainably increase productivity caused its fall, because it was a failure of all civilisations until that originating in what we now call modern western Europe. The fall of Roman civilization clearly had a whole range of interconnected causes. One was increasing monetary problems largely a result of the dependence on coin for economic exchange, though the first steps towards managing the money supply were taken, with the manipulation of the coinage from the third century. In the short term, this appears to have contributed significantly to the demise of banking as a profession. Another enduring innovation was the introduction of the national budget by Diocletian. Nevertheless, the last western emperor died in 476 AD, although Roman traditions were continued, and in the legal field at least continued to develop, in the East under Byzantine rule. In the west, political rule fragmented, and social institutions took on a wide variety of less sophisticated forms. Nevertheless, once political stability started to return, cumulative developments started to take place again, combining elements of the ancient heritage with a whole range of novelties. It is to this potent mix that we now turn.

2.2 Medieval Europe

Post-Roman Western Europe became, more than anything else, fragmented. So it is not possible to be as specific about institutional arrangements as it was with the ancient world, where Athens to a large extent, and then Rome totally,

dominated the region, at least as far as institutional sophistication is concerned. Nevertheless, it is possible to make some general statements, as well as identify some key developments.

While the barbarian conquerors of the western empire, not surprisingly, appear to have liked the idea of taking up the lifestyle of the wealthy Romans, they were oblivious to the means necessary to sustain it. So the government institutions, and the services they provided, quickly fell into disrepair. Interconnections were lost, and communities turned in on themselves to sustain themselves. Gaps opened up physically between communities. The reasons for this are not completely clear, but some evidence suggests that the last centuries of antiquity and the first of the middle ages were particularly cold and wet, and alongside this plague was intermittently common from 180 CE to the mid sixth century (Lopez 1976, 12). Taken together with the breakdown of the *pax romana* and the return of warfare to common experience, sometimes associated with the movements of people such as the Anglo-Saxon invasion of the Britain, all this suggests that, although there is no direct written evidence, there was probably a considerable decline in population.

The paucity of evidence makes identifying the point at which the decline was reversed impossible, even if there were to be one such time. Nevertheless, by the tenth century the demographic tide had turned and population was starting to grow again, although it took until the fourteenth century to occupy many of the wide uninhabited stretches in a way that made communication and organisation practicable, as well as increasing the food supply (Lopez 1976, 29). One indication of a general improvement, albeit far from signifying much in the way of day-to-day improvements, is that evidence of large-scale famines become rare by the early thirteenth century (Lopez 1976, 37).

An indication that social structures had changed, and that changes made a difference, is evidence of the widespread diffusion of water mills. Lopez

identifies two factors to explain why, although known about in the ancient world, widespread utilisation had to wait. First, that there were no longer slaves available to do the work, but landlords still had sufficient power to compel their tenants to send their grist to the mill. Then, for estates lacking a usable water source, the windmill became available. Not documented at all in antiquity, by the early twelfth century its use was widespread. Even by the late eleventh century, England is recorded in the Domesday Book as having five thousand mills. Another improvement in productivity came through the introduction of the heavy plough. Again its origins are mysterious, and it could have been known in Roman times, but its use only became widespread during the middle ages (Lopez 1976, 43-44).

The problematic nature of generalisation comes to the fore again when considering medieval trade. On the day-to-day level for most people there was little trade. This is perhaps exemplified by the fact that archaeological evidence suggests that money disappeared from use in England for two centuries, from shortly after the Romans left at the beginning of the fifth century, until the seventh century. Even though coins were available on the European continent, there is only evidence that they started to be imported shortly before local production recommenced (Davies 1994, 117). Nevertheless, Mediterranean trade was sustained, to a large extent by easterners such as Jews, Arabs and Greeks (Lopez 1976, 60). By the tenth century it is apparent that the Italian cities of Venice and Amalfi had become important hubs. Interestingly, neither city had any Roman past. An additional novelty was that master craftsmen appeared in the upper classes (Lopez 1976, 62). So a new type of polity had started to appear, one that focused on making a living from craftsmanship and particularly from trading. It began to prove increasingly successful. By the early fourteenth century Venice, Milan, Florence and Genoa all had populations in excess of 100,000, twice the size of any city in any other part of Europe (Lopez 1976, 93).

Societies were clearly more flexible than in ancient times, though sustainable economic growth remained elusive. Significant innovations were introduced, and one in particular, the printing press, will be considered in some detail. Despite the astonishing variety that characterised the middle ages in Europe, by concentrating on the most commercially successful regions of their time it is possible to sketch out the further development of embryonic institutional commitment and credit mechanisms.

2.2.1 Commitment

Post-Roman Europe's fragmentation encompassed the legal arena. The northern Germanic tribes had their own, albeit unsophisticated, legal principles, which they wished to maintain, and southern Europe maintained a good deal of continuity with Roman law, at least an unsophisticated derivation of it. In practice there was a large area of overlap, and together with fragmentation and conflict between communities, the result was that legal practices evolved on divergent paths.

Variety in contractual forms was as marked as in every other aspect of medieval life, although there were to some extent moves towards standardisation. Central to this was the growth of the legal profession, which was very much a pan-European phenomena. Nevertheless, by the end of the period two distinct traditions had become established, the common law tradition in England, and the Roman law tradition, which later evolved into the civil law tradition, in the rest of Europe. However, divergence very much preceded convergence, and the interplay of the increasing demand for greater sophistication of the wide variety of local laws rooted in tribal customs, and the rediscovery, and sometimes reinvention, of Roman law provided a heady brew.

As in so many ways, Italy with its direct Roman heritage led the way. The tenth century saw the introduction of the *contratto di commenda*, also known in Venice as the *collegantia*. It was used in trading, and had characteristics of both

loans and partnerships. It incorporated a distinct division of labour between the provider of capital and the travelling merchant. Any losses were borne fully by the capital provider, but profits would be shared: one quarter by the traveller and three quarters by the capital provider. The capital provider could not be held directly accountable for the activities of the traveller, and both partners were able to enter into other contracts as they wished. Capital providers could invest relatively small amounts along with others, with the rewards being shared in proportion to their investment (Cipolla 1994, 161).

In time practices changed, and the *commenda* became outdated. Beginning in the latter part of the thirteenth century, and accelerating during the second half of the fourteenth, businessmen operating through agents superseded itinerant merchants. In its place emerged the *compagnia*. By the early twelfth century Venice already had several such companies, but they became more common in the inland cities that combined production with trade (Cipolla 1994, 162).

Italian jurists in the twelfth and thirteenth centuries 'found' the concept of partnership in Roman law, with partners sharing risk and profits, and insisted that a commercial society was a 'brotherhood' of equals. Most were short lived and small, but some could be ongoing with a membership reaching well into double figures. By the early fourteenth century greater sophistication was needed. One example is the Buonsignori of Sienna, whose junior partners petitioned their republic to intercede with the papal court in establishing the principle that no partner be liable beyond his share in the company. Rejected initially, by 1310 it was accepted. Meanwhile, in 1301 the *Calimala*, the Florentine wool guild, had provided that the creditor of a bankrupt partner could only take their shares, not pursue the firm itself (Mundy 1991, 100-2). Recognisably modern forms of legal association were slowly beginning to emerge.

From the 1160s, the *montes* created by governments, and shares sold in such revenue streams as harbour duties and state salt monopolies, can be considered as public private partnerships. Essentially, these investments were limited liability vehicles, as investors in public concerns risked no more than they had invested. Sometimes governments intervened more aggressively. In 1283 the Venetian government ordered its merchants in Alexandria to combine to corner the pepper and cotton trades (Mundy 1991, 103).

While many developments were shared, polities jealously guarded their legal autonomy, but the growth of international trade created a demand for a common system. From the twelfth to the fifteenth centuries a customary commercial law developed, known as the *lex mercatoria*, or law merchant. Its source was the practical needs of merchants, and covered a whole range of issues from corporation to credit, through trade fairs, banking operations and insurance. Merchants had their own courts, and were judged by their peers (Caenegem 1992, 83-84).

Given the medieval conception of law, as something having an independent existence, and being essentially unchanging, and not subject to human manipulation, the distinction was primarily one of procedure over substance (Baker 1979). A modern translation of the earliest extant treatise on *lex mercatoria*, from *The Little Red Book of Bristol*, circa 1280, identifies three differences from the English common law. First in being faster in making a determination, second in a broader interpretation of pledging that included responsibility for resulting damages and expenses in addition to the specific debt, and thirdly a difference in evidence and proof that placed the burden of proof for informal debts on the plaintiff, but required the defendant to address this directly (Teetor 1962).

While, in general, courts were local in their jurisdiction, *lex mercatoria*, was regarded as universal. In the treatise already cited, although found in a collection of documents in Bristol, no mention is made of Bristol, while it

does mention London, Boston, York and Paris. Whether or not it was particularly relevant to Bristol, the author clearly expected the principles to be commonly accepted in major cities across national boundaries. In the fullness of time, as part of the process of nation building, *lex mercatoria* was incorporated into national legal systems, and in effect ceased to constitute a common international commercial legal framework. Nevertheless, in its time it provided an effective institutional foundation for the development of commerce, and because used and administered by merchants, was able to evolve in step with other developments in commercial practice.

England's distinctive economic trajectory began to emerge in the thirteenth century, because rather than fall back on Roman tradition as a framework for providing a consistent legal framework, the common law tradition continued to evolve. At least one significant cause of this was the creation of a single national customary law by Henry II (1154-89) (Caenegem 1992, 35). Influenced by the developing Roman scholarly tradition, this also included the recognition of the importance of written records, and a complete set of 'rolls' of the royal courts exists from the twelfth century onwards (Caenegem 1992, 96). The medieval recovery of Roman law was not sufficiently well developed by this stage to provide a comprehensive written source, so the precedents recorded by the royal courts became the primary source of law in England, while elsewhere in Europe, where written sources of customary law did not exist to the same extent, there was nothing to compete with the written authority of recovered Roman law once it became available.

Through the thirteenth century the shift began from a feudal notion of property, to what could be labelled a modern one centred on individual ownership. With a rising population, the demand for land made it more valuable, and two English statutes, Merton in 1235 and Westminster in 1285, permitted a manorial lord to enclose wasteland, but only so long as sufficient land was left for tenants. Most significant was the ability to transfer the ownership of land. The feudal perspective was that several people had

jurisdiction over any piece of land, namely the hierarchy of peasants, lords and kings into which everyone fitted, and everyone had their own rights. Transfer of land could only happen to the extent that one tenant could return his rights to his lord, who could pass them on to another, or a tenant could create another feudal layer by granting part of his rights to another, an action known as sub-infeudation. However, the statute *Quia emptores* in 1290 forbade sub-infeudation, as well as recognising what had become common practice recognised by the king's court by then, namely transfer of land between free men without the lord's permission. By the end of the thirteenth century the concept of the freeman, as one whose obligations were strictly defined, who could transfer his land holding and was answerable only to the king's court, had been established. This is not to say that most men were regarded as freemen, only that the concept had become accepted. North and Thomas suggest that the significance even of that is difficult to overestimate. In taking this step, England set out on a trajectory differentiating itself from the rest of Europe, and ultimately the world (North and Thomas 1973, 63-64). The next chapter sets out in detail the medieval origins and subsequent development of contract law in England. But first, developments in medieval finance will be set out.

2.2.2 Credit

Without a central government to direct commerce, as had been the case to a greater or lesser degree in the ancient world, in the medieval maelstrom a decentralised mechanism was necessary to organise the flow of goods. While by no means a new phenomenon, what was in many ways new was the relationship between the ruler and the merchant. Central to these developments was perhaps the ultimate physical embodiment of the idea of the free market, the medieval fair. Rather than simply emerging spontaneously at a convenient location, however, a fair was established by a charter. And charters described the public goods that would be provided to market participants. These included safe conduct for merchants, security for their property, and special courts to resolve disputes. Moreover, there is also

empirical evidence that earlier more local economic activity supports a principle of 'no mint, no market' (Hunt and Murray 1999, 26). So currency is also a public good necessary for a functioning market. The rise of the fair was rapid, even by modern standards. In England, the first 75 years of the thirteenth century saw the granting of 2200 separate charters for fairs and markets (North and Thomas 1973, 50), where the doomsday survey of 1086 had identified around 3000 communities (North and Thomas 1973, 44). Localised famines, which had been widespread before 1200, became less common during the thirteenth century (North and Thomas 1973, 53). Only in the decade 1307-17 did famine spread again throughout Europe, and recurred intermittently, being followed thirty years later by the start of the Black Death.

During the thirteenth century, the first genuinely international fairs developed in the Champagne region of north central France. This lay across the trade routes connecting the northern markets in England and Flanders with the Italian and Mediterranean markets in the south. Between January and October there were six fairs. These were originally religious festivals, but the Counts of Champagne courted merchants by offering safe conducts and special privileges. Northern merchants brought primarily cloth, but also specie, and exchanged it for luxury goods made in the Mediterranean region, or imported from further east.

In conjunction with the growth of the international exchange of goods, there developed payment systems, and in particular credit instruments known as 'fair letters' to facilitate the trade in goods. To a large extent these obviated the need for large shipments of specie. They were not negotiable in the modern sense, but could be transferred from one person to another (Hunt and Murray 1999, 29). Nevertheless, they offered a flexible way of committing to make payments in a specific currency and time agreed between two parties.

While overland trade, including river transport, never went away, sea borne trade became increasingly important. Alongside this, merchants, and

particularly Italian merchants, started to establish branch offices in major trading centres outside their own cities, particularly in the northern ports such as Bruges and London. So by the end of the thirteenth century, the champagne fairs were becoming less significant.

A more general instrument than the fair letter also developed, the bill of exchange. One early version dates from late twelfth century Genoa. The transaction involved four parties, the lender and borrower in the town of issue, and their representative and correspondent respectively, who would actually exchange the cash, in the town of repayment. Payment was scheduled for a future date, and normally was in a different currency to that in use where the transaction was agreed, because this allowed the lender to include a margin, which the usury laws of the time allowed, as the currency fluctuations meant that they bore an element of risk. In general, the payment due date reflected the time normally expected for a cargo to travel to the place where payment would take place, and was known as 'usance'.

A phenomenon recently dubbed the 'super company' flourished briefly towards the end of the thirteenth century, then disappeared in the 1340s. They were trading organisations that covered the whole range of goods traded between northern and southern Europe. Their demise was probably due to their inability to cope with fluctuations in the grain supply following a concentration of extreme weather conditions. Not long afterwards the Black Death swept across Western Europe, with devastating human consequences, but leaving behind an environment ripe for organisational innovation.

Two shifts started to become apparent in subsequent decades. One was the end of Italian dominance in international business, with the rise for example of the Fuggers in southern Germany. The second was the increasing involvement of large-scale merchants, including the Fuggers as well as Italians such as the Medici, in production as well as trade, albeit not on such a

concentrated scale as had occurred with the super-companies (Hunt and Murray 1999, 169).

The growth in trade, and associated moves towards specialisation, suggests and explains a growing productive capacity. Some of this growth was in luxury goods, but much of it was consumed in military activity. Either way, the major beneficiaries were the rich and powerful. But the other beneficiaries were town dwellers who produced and traded the growth products. Where faced with relatively weak feudal rulers, in areas such as Germany and the Low Countries, they were often able to organise themselves and manage their own affairs to a considerable extent. Feudalism never really took off in northern Italy, where the landowners tended to move into the towns at a relatively early period. One effect of this was that Italian city-states often had ruling elites who, because of their heritage as large landowners, were geographically acquisitive, leading to a continuing series of wars. Partly because they lacked a large agricultural population able to provide men for military service, and partly because of the wealth created through their production and trade, warfare became increasingly capital intensive. Rather than conscripting feudal underlings, foreign mercenaries were engaged. And the combination of competition and availability of capital from commerce led to rapid developments in military technology, spurred considerably by the growing use of gunpowder.

While at first the funds necessary to fight increasingly expensive wars were extracted by taxation and 'loans,' on an irregular basis, the continuing and increasing need led to a more systematic process. Although many lenders never expected to be repaid, loans were often only provided in return for concessions such as trade monopolies. Particularly given the international nature of much commerce, merchants could rarely be compelled to supply funds, so a degree of mutual benefit was often present, unlike the zero sum outcomes likely in monolithic autocratic polities. Nevertheless, as early as the mid thirteenth century Venice and Genoa, followed within a century by

Florence and other Italian cities, had regularized their public debt. Known as the *monte*, because of its mountain-like size, this was accumulated by the cities, but paid lenders a regular low rate of interest. For example, at the turn of the fifteenth century Venice paid four percent. The debt was transferable, and traded at rates that varied with commercial rates and the relative financial strength of the city (Hunt and Murray 1999, 207). Further north, to meet their own needs, and the demands of the relevant feudal ruler, cities tended to prefer to issue annuities, where they agreed to pay a fixed sum to particular individual, and possibly their next generation, in return for a lump sum payment (Hunt and Murray 1999, 208). While most of these public funds went directly or indirectly into sustaining the lifestyles of the elites and in unproductive military activity, some went into capital projects, such as river widening and voyages of exploration.

While these to some extent mutually beneficial approaches developed in the politically fractured Italian and German regions, in more autocratic regions, particularly France, another alternative developed: 'reforming' the coinage. Rulers benefited from this in two ways, first through seigniorage, the charge levied for the actual exchange of coins, but also from the precious metal retained when the newly minted coins were exchanged for those with a higher content of precious metal, though nominally retaining the same value. While within the state's boundaries this could be maintained, and there could be some benefit to trade from an increase in the money supply, exchange rates would adjust correspondingly, making people poorer by international comparison (Cipolla 1994, 172).

During the sixteenth century the role of Italians declined, and the Low Countries became the commercial heart of Europe. Commercial banking developed in both volume and expertise as the middle ages drew to a close, to the extent that rates of interest dropped from twenty to thirty percent in 1500, through nine to twelve percent in 1550, to three percent or less during the seventeenth century (North and Thomas 1973, 142).

2.2.3 Consequences

Medieval Europe was in some ways more hierarchical than Greek or Roman society, though most of those were in theory rather than in practice. The great difference was in the variety of the medieval world compared with the homogeneity of the classical world, certainly under the Roman Empire. And this variety was constantly interacting, increasingly so as the age progressed. In this environment opportunities for novelty became commonplace. Alongside the majority of innovations that were borrowed from elsewhere, there were some that were genuinely new.

A relatively well-documented series of technical developments can be set out. From the sixth century use of the water mill spread, then from the seventh, particularly in the North, the heavy plough. The eighth century saw the spread of the crop rotation system, and the ninth the horseshoe and new methods of harnessing draft animals (Cipolla 1994, 138). Windmills spread through the twelfth century, and the compass developed from a primitive needle in a bowl to a self-contained instrument from the end of that century right through the next. At the beginning of the fourteenth century came the spinning wheel, spectacles, clocks, firearms and canal locks. A realisation was dawning that technological change was possible (Cipolla 1994, 143-5).

In transport, significant improvements occurred, particularly in the fifteenth century with the development of the full rigged ship. Within a century, ships were developed to a size at which they largely remained until the nineteenth century. This made possible significant improvements in trade, but also provided the technical foundation on which the European voyages of discovery were made (Cipolla 1994, 145-8).

One innovation more than any other contributed to the end of the medieval world, and for the first time in history, though disputed from time to time in the past, we know the name of its inventor. This is no coincidence, for the invention changed the very nature of human knowledge. Around 1400, in

Mainz, Johann Gutenberg was born into a rich merchant family. Family names were not as fixed as they are now, often being taken from the name of the family property, and the name by which we know him was taken from the property he was born in and later occupied. His father was born around fifty years earlier, just after the Black Death. His mother was the daughter of a shopkeeper, albeit one whose grandfather had been governor of the fortress of Mainz, until 1332 when a dispute between the city and emperor erupted. The city lost; with lasting social consequences, if not so much economic ones, for the families of those city patricians on whom primary blame fell. Apparently, this somewhat mixed social pedigree would have limited Gutenberg's prospects in his hometown, at least compared with his father (Kapr 1996, 30-31).

One particular formal association held by Gutenberg senior, but denied to the son, was being a companion of the mint. There is evidence that other holders of this office resided for some time at the same dwelling as the young Gutenberg. So although denied the opportunity of exercising the office, he could well have perfected many of the skills needed in coin production (Kapr 1996, 49). A Combination of family connections and greater opportunities offered by a larger town were probably behind Gutenberg's move to Strasbourg sometime before 1434 (Kapr 1996, 67). Here we find the first direct evidence of his entrepreneurial spirit. Around 1438 he entered into a partnership with three other men to manufacture ornate convex mirrors prized by pilgrims for catching the power of religious relics shown once every seven years at Achen (Kapr 1996, 72). Gutenberg clearly had the ideas and the know-how, but needed the capital and labour of his partners to realize the opportunity. All expected to become significantly wealthier.

Fortunately for us, the partners' plans went awry. The pilgrimage expected in 1439 was in fact only to take place in 1440, and in the meantime one of the partners succumbed to the plague. The deceased partner's brother proposed that he and an older brother take his place in the partnership. Wishing to

maintain secrecy Gutenberg refused, and was taken to court, and well-attested copies of the records still exist (Kapř 1996, 75).

What is in many ways more interesting, is that the court records say that when the miscalculation of dates was realised, the men entered into a further partnership to pursue activities already being developed by Gutenberg, but are only described in the most general terms as, for example, 'art and adventure', or 'art and invention'. Nevertheless, the evidence from the legal records can be interpreted to suggest that the three key stages of print production, namely typesetting, typesetting and pressing, were undertaken at the three partners' own residences. Although more complicated, Gutenberg may have preferred the secrecy this provided, as well as cutting down on the costs necessary for a dedicated workshop (Kapř 1996, 86).

The last record of Gutenberg's presence in Strasbourg is for 1444, the year his five-year partnership was due to end. The next reliable document has him back in Mainz, albeit only in 1448. Part of the reason for Gutenberg's departure may be the appearance around Strasbourg at that time of a band of unemployed mercenaries known as the Armagnacs, recently put out of work by a truce between the French and English.

It is likely that Gutenberg did produce his earliest books as Strasbourg, but the technology was still rapidly developing, and there is no indisputable evidence that this was the case (Kapř 1996, 99). Gutenberg had returned to Mainz by 1448, and began printing in the building where he had largely grown up, now owned by his widowed brother-in-law. Nevertheless, to produce a substantial book, in 1450 he entered into another partnership, with Johann Fust, a merchant and goldsmith. Fust's capital allowed the renting and equipping of a workshop with initially four and later six presses, with the first tranche of funding being secured against the equipment being assembled. Subsequently, Fust supplied additional capital, as around 20 workmen would need to be recruited, trained, and paid. Supplies of ink and paper needed to be

purchased. Also thousands of types cast, a task estimated as taking six months. All this before production could actually begin. Production would probably have taken two years, and none of the estimated 180 finished copies of Gutenberg's now famous 42 line Bible could be sold until then (Kapr 1996, 158-67). Nevertheless, the evidence of a Bishop's letter dated 12 March 1455 suggests that the project had proved a success, and buyers were quickly found. People as far apart as Vienna and Rome were aware of this new invention within no more than a few weeks.

However, while the project was a success, the partnership was not. Again, we benefit from Gutenberg's misfortune, because court records are still in existence, albeit not in relation to the original action, but from one several months later in November 1455. Quite what the underlying motivation behind the dispute was is probably impossible to know, but Fust sued Gutenberg for improperly using funds from their partnership, and for interest on money he had needed to borrow to invest in the project. Despite the completion of the Bibles, which rough calculations based on reliable evidence suggest should have been profitable, Gutenberg cannot have yet received payment and, admitting the charge to some extent, was unable to repay the money. So Fust took sole possession of the major book-printing workshop (Kapr 1996, 184). Kapr suggests that, on the evidence of separately set but identical in content indulgence documents, which the Church used to raise money in return for remitting sins, Gutenberg was still using his original workshop to print these. Given the likelihood that he was able to price them at the same level as a hand-produced version, but more accurately and more rapidly produced in large quantities, these would have been very lucrative. Gutenberg may have needed the funds for continuing with his Bible printing, in addition to those supplied by Fust, which though profitable required considerable capital, some from the profits of indulgences. The identical, except in setting, indulgence forms suggest that Fust switched to the more immediately profitable activity of producing these, once he gained control of the workshop (Kapr 1996, 196). While Gutenberg never left his name on his

work, Fust and his adopted son, and later son-in-law, Schöffer, applied their imprint to editions produced after the dissolution of the original partnership.

While the end of the partnership with Fust was clearly a blow, it would not have left Gutenberg in any immediate financial difficulty. Nevertheless, he continued with his work, and by 1457 he had defaulted on the interest payment for a loan he had taken out in Strasbourg (Kapř 1996, 221). Between 1458 and 1460 another edition of the Bible was produced at Bamberg. The likelihood is that this was a commission by the Prince-Bishop, who paid for the entire costs of the 80 or so copies for distribution around the diocese. It is likely that Gutenberg was involved in this project, but that it was primarily managed by one of his former assistants (Kapř 1996, 224-5). Not long after the Bible was finished, the type, workshop and presses were sold to the secretary to the Prince Bishop, who continued to produce small popular works, many in German. The quality of these suggests that the most skilled craftsmen returned to Mainz after the sale (Kapř 1996, 225-6).

As early as 1458, books also started to be produced in Strasbourg, under the management of another of Gutenberg's assistants. Then in 1462, in a dispute between two claimants to the Archbishopric of Mainz, one who supported local autonomy from Rome, and another appointed by Rome on the excommunication of the first, resulted in an attack on the city, and the expulsion of many of the inhabitants, almost certainly including Gutenberg and his associates. Likely exiled to Eltville, where he had relatives, there is evidence Gutenberg was involved with a printing works there. While Gutenberg was subsequently allowed to return, with some honour, to Mainz for his last few years, in the vicinity of which he died in 1468, his associates dispersed together with their skills across Europe. By 1500, less than fifty years after Gutenberg's first major production, it is estimated that at some 255 locations more than 30,000 publications had been produced, in a total of over 20 million copies (Kapř 1996, 285).

The details sketched of Gutenberg's innovation demonstrate two things. First, and perhaps of less importance, is the role of legal records in providing data on which much of the description of developments is based. Secondly, is the nature of the innovative process, and the fundamental role of institutional commitment and credit mechanisms in enabling it to function.

By modern business standards production of the first printed books was not a large project, but it was clearly too big for one person to undertake themselves. Formal contracts were clearly important in motivating both the active and investing partners to supply their time and resources. Fust's claim to have borrowed the money he invested may have been simply a legal fiction necessary to counter the defence that he was a usurer (Kapf 1996, 175). But it is quite possible that as an active merchant he did not have the money to hand. Either way, loaning money was clearly not an unusual practice, even if it was normally for working capital rather than capital investment as in this case, and interest charges were recognised as facts of life, if not as morally or legally acceptable. Fledgling credit markets existed. Cipolla points out the significance, even though existing on a relatively small scale during the middle ages, of the institutional mechanisms for transferring funds, temporarily, from savers to innovators and producers (Cipolla 1994, 164).

The limitations in the institutional environment supporting contractual commitment were also apparent. The Gutenberg case demonstrates how on two occasions, what should from a contemporary perspective be relatively straightforward arrangements became the subject of expensive litigation and in the second case a damaging outcome. It is possible that effective intellectual property legislation may have helped Gutenberg but this was only just starting to be developed. Venice granted specific patents from 1416, and promulgated a formal patent law in 1474 (Hunt and Murray 1999, 154). While the evidence suggests that Gutenberg was not motivated by material gain for its own sake, he recognised the need for commercial success to enable his work to proceed. The dispute with Fust may have arisen because while

Gutenberg saw this as simply an enabler, he realised that for Fust it was the only objective, and if indulgences were so profitable he would see no point in producing books. The idea that continuing innovation was the only way to sustain profitability in the long term was clearly not yet apparent to businessmen.

Further south, Italy's preeminence in medieval commerce was about to decline. Between 1494 and 1538 it became the battleground for international conflicts between the increasingly centralised nations of Spain and France, as well as trade-enriched, though remaining decentralised, Germany. With the war and its disruption came famine and disease, creating a downward spiral. The second half of the sixteenth century saw something of a recovery. But the next century saw Italian products increasingly losing out to English, Dutch and French ones. Italian quality remained pre-eminent, but not enough to bridge the price gap that developed. The Italian insistence on traditional methods and maintaining quality, enforced by the guilds, proved misguided. Its great cities lost their vitality, and its most important exports became agricultural and semi-finished goods (Cipolla 1994, 241-7).

While many of the legal foundations had been developed in Italy, they applied almost exclusively to trade, rather than to production. And while the guilds were very successful in maintaining the quality of production, they also prevented innovation in production. And it is innovation in production that is the locomotive for sustainable economic growth. The Hanseatic League in Germany and the northern Netherlands republics also had many of the building blocks in place, but partly because of the disruptions caused from time to time by military action, and partly because the individual economic entities were relatively small, and relied on ongoing personal relationships which significant innovation could disrupt, the right sort of innovation failed to take place here either. Enduring commercial relationships needed to move from the personal to the impersonal.

The Dutch did rise to commercial primacy, and have been cited as the first nation to experience sustained growth (North and Thomas 1973, 132). But this was the 'fluke' of being the last to start on the well-established path of a commercial upswing. Inevitably, this was followed by a downturn, but whereas before genuine sustainable growth was kindled this would have been an absolute decline, Holland in fact benefited from the effect of sustained English growth, to the extent that it only experienced a relative rather than an absolute downturn after its ascendancy. They were caught in the slipstream. This is not to take too much away from the Dutch achievement; before England attained their ascendancy just before the start of the eighteenth century the English were very aware that they had a great deal to learn from the more advanced Dutch (Cipolla 1994, 259).

By 1700 in England however, the pieces were in place. While in most cases the pieces were embryonic, a symbiotic dialogue developed between the needs of business and the formal legal foundation necessary to enable effective organisational and credit relationships. Over the next century and a half, by the end of the 1850s, the structures we recognise now as necessary for modern business were firmly in place. It is now time to turn to describing the nature of the relationships made possible for the first time at that time, and sketch the processes by which they developed.

Chapter Three

INDUSTRIAL REVOLUTION IN ENGLAND 1700 - 1850

In some ways more precise historical dates could have been used for the title of this chapter, namely 1689-1856, because that is time from the 'Glorious Revolution' until the introduction of the Limited Liability Act. And those were in many ways the defining events of their age, at least for the issues described here. However, the level of precision suggested would be spurious, because these events were simply steps on the ladder. Neither was wholly unexpected, although neither was a forgone conclusion either.

By setting the boundaries as 1700 and 1850, within the symbolic defining events, the focus is on the business environment as it existed and evolved, rather than the major changes at the beginning or the end. The economic, legal and political structures that, since the time of the fall of the Berlin Wall in 1989, define large-scale social relationships around the world were already in place in England by 1700, and by the end of the 1850s a recognisably modern framework of corporate governance had developed there too.

One economic historian has said that there is a case for arguing that the division between ancient and modern occurred in 1709, when in England one Abraham Darby first successfully smelted iron with coke, and introduced the concept of self-sustaining growth to the world (Earl quoted in (Landes 1993, 132)). While smelting iron with coke was a considerable technological advance, in itself it is clearly difficult to even to begin to justify the claim from the perspective of technical innovation. What does make the claim plausible is that Darby founded a business that, albeit as an indistinguishable element, is still in existence to the current day. And in the course of its almost three hundred year independent history, the company also produced the first iron wheel, the first iron rail, the first iron bridge and the first steam powered

railway locomotive. Clearly it benefited from its access to the materials necessary for iron making, but as a self-sustaining organisation the firm took a form and existed in an environment never experienced before in human history.

To help understand the distinctiveness of the organisations and environment existing in England around 1700 it is instructive to compare Darby's experience with that of a contemporary born four years later than him, but in Germany. He was also what would today be called the developer of an innovative material; Johann Böttger was the first European to make porcelain.

Böttger, of course, although working independently, was not the first to make porcelain. The Chinese developed it in a gradual process around the sixth century (Gleeson 1998, 40). But Darby's distinctive originality was not so much in the material, novel though it was, but in his market. In part the value of smelting iron with coke was that it was as an alternative to charcoal, and in England the coal used as the raw material for coke was available in substantial amounts, whereas the wood used as the raw material for charcoal was in increasingly short supply. Even so, the iron smelted using coke was different from that smelted using charcoal. The latter was less brittle, to the extent that the former was difficult to forge, and not in demand by the forge-masters who constituted the primary market for raw iron. However, while less suited to forging, coke-smelted iron was better for casting.

Darby had his insight no doubt partly because as a youth he was apprenticed to a malt-mill maker, someone who produced small brass machines used to process malt for use in brewing, and he subsequently set up his own brass company, so would have understood the markets for brass (Thomas 1999, 4). It was also probably because he was born close to the home of 'Dud' Dudley, a landowner who had claimed to be able to smelt iron with coke towards the end of the seventeenth century, though not in commercial quantities (Thomas 1999, 12). The properties of cast iron meant that it would be an effective

substitute for brass. However, whereas porcelain pots, and even Gutenberg's printed books, were clearly aimed at the wealthy, the primary benefit of cast iron over brass, as it would clearly be less aesthetically pleasing, was its cost.

Coke-smelted iron was, for the first time, an innovation aimed purely and simply at the mass market of working people. The founding of Darby's company at Coalbrookdale signalled the start of what would now be called the democratisation that enabled the industrial revolution, and what we now call sustainable economic growth.

Darby's and Böttger's genius extended far beyond their technical ingenuity. It encompassed a significant skill in building relationships and managing what were, for their time, complex long-term capital-intensive development projects, as we have already seen for Gutenberg. The environments and markets, and the specific approaches these required, were however worlds apart, worlds we will briefly compare, because they exemplify the distinction between the modern and pre-modern worlds.

Böttger's father, a mint-worker, died soon after his birth, but his mother remarried a well-educated widower who was an architect and engineer. This enabled Böttger to benefit from an education, and develop his early interest in chemistry, becoming apprenticed to a leading apothecary in Berlin (Gleeson 1998, 13). His curiosity brought him into contact with several individuals who claimed to either have or be looking for the 'philosopher's stone' of alchemic legend, which transmuted other metals to gold. The benefit of such a discovery was not lost on the rulers of the day, and several German princes supported some of these people (Gleeson 1998, 14). It is impossible to know precisely what his motivation was, but Böttger contrived a display in which he appeared to be able to transmute silver into gold. Although requesting secrecy from the few people to whom he made the demonstration, news quickly spread. He was summoned to appear before the King of Prussia, but deciding his luck was about to run out, he absconded. Unfortunately his perceived

value was such that he was pursued, although because he had fled to the neighbouring kingdom of Saxony his apprehension by the Prussians required the acquiescence of the Saxon authorities.

Curious to the reason why the King of Prussia was so interested in a common subject, the Saxon authorities made their own enquiries and, when they found out, informed their own king. He decided to keep Böttger for himself (Gleeson 1998, 18-22). Despite, not surprisingly, failing to demonstrate the production of gold, promise of its imminence was enough for the credulous king to appoint him master of his mint. In the intervening time, however, Böttger had met one of the King's councillors, Ehrenfried von Tschirnhaus, who among other things was responsible for locating promising mineral deposits in the kingdom. One of his interests was porcelain, at the time imported in considerable amounts and at huge cost from the orient. Tschirnhaus became a frequent visitor to what was effectively Böttger's prison.

After three years, by the spring of 1705, it was becoming clear to the king that Böttger's claims were spurious. However, he was concerned that he would lose considerable face if he executed Böttger, because he had invested so much in him. At Tschirnhaus's suggestion, his remit was widened to include the production of porcelain (Gleeson 1998, 44). He was moved to larger premises and supplied with samples of all the clays produced in the kingdom. Böttger started to attack the problem systematically, and realised from his previous experiments that far higher temperatures than those used in conventional pottery or glass production were required. Political events soon intervened, however. An unsuccessful war with Sweden was leaving Saxony exposed to attack, and Böttger was moved along with the King's most valuable possessions to a remote castle prison. After a year, events calmed down, and Böttger was able to resume work, in another location with a bigger and more powerful furnace built for him. By 1708 he had produced samples

of porcelain, and on 28 March 1709 wrote to the king declaring himself ready to go into full-scale production.

Darby's path to his 1709 success in England was somewhat more mundane. After completing his apprenticeship in Birmingham he moved to Bristol, in 1698. Like his parents and his apprentice-master he was a member of the Society of Friends, more commonly known as Quakers. The year after his move to Bristol, Darby married another Quaker. He founded his first company, the Bristol Brass Company, around 1702 to make malt mills, with three partners. He seems to have foreshadowed his later work by using coal rather than charcoal for the brass making furnaces. Not long afterwards he established another enterprise, initially alone but then with two other partners, to cast iron pots. He had used techniques developed in Holland to cast brass in sand, rather than clay or loam, and wanted to utilise the same technique for casting iron, but it was proving problematic for the Dutch brass-founders he employed to make headway. Working on the problem himself with a talented assistant he was soon successful, and in 1707 successfully petitioned for a Patent giving him legal protection for fourteen years exclusive use (Thomas 1999, 9). While in Bristol he also came to know Thomas Goldney II, also a Quaker, who was a prosperous merchant.

By 1708 Darby had become aware of a disused furnace, which had been out of action since being partly destroyed in an explosion and then its operators subsequent move to Russia several years before, presumably as part of Peter the Great's Europeanisation project. It was in Coalbrookdale, further up the River Severn, which reaches the sea in Bristol; and it offered a considerable number of advantages, being close to sources of iron ore and coal, as well as easily accessible by river to Bristol and the sea. In September 1708 he took over a lease previously established until 1717. Initially restarting the furnace using the tried and tested charcoal fuel, he started to experiment with coal and coke, and although exact dates are uncertain, it is likely that he was successful

fairly quickly, in 1709, but needed another few years to hone the technique to his own satisfaction.

Böttger too was still some distance away from a reliable production process, but his captor still insisted he would not be free until he started producing gold, as he had originally promised to do. Böttger's spirits cannot have been improved by the news that another alchemist, employed by the King of Prussia following Böttger's escape, had been publicly hanged as a fraudster, accompanied with much golden decoration. Nevertheless, the King of Saxony cautiously provided more backing to Böttger's porcelain making endeavours. One major concern was secrecy, and for this reason the key stages of the production process, namely mixing, modelling, firing and decorating, would be carried out at separate locations.

The King was still reluctant to invest directly, partly because he was involved in another war, and had the opportunity to invest advertised throughout the kingdom, but to no avail. Undaunted, the King proceeded, albeit cautiously. Key managers were appointed, and recruitment of craftsmen started. The recruiting of managers proved straightforward, presumably there were no shortage of courtiers keen to please the court and seeking opportunities for enrichment; but finding craftsmen proved more difficult. Despite the novelty and interest of the project, presumably a regular wage could be problematic, given the people responsible for the enterprise. When limited sales did commence, the number of purchases proved very disappointing.

A move to larger premises was made, and production started to gain pace. Finance however continued to be a continual worry. Funding promised by the King was erratic, and was not improved by fact that the director he appointed was redirecting considerable amounts to his own pocket, as were many others. Perhaps less dishonest, but at least as damaging, was the King's appropriation of a considerable portion of output for his own use, either directly or as gifts, without regarding it as necessary to pay anything for them.

In 1711 a commission was established to investigate the obvious problems, and Böttger was able to successfully make the case for proper investment and a reorganisation of the management. Meanwhile, numerous technical problems were still being addressed, and a skilled workforce recruited and trained.

By 1713 white porcelain was being produced in commercial quantities. Böttger's achievement can be grasped to some extent from the fact that no one was able to improve on his basic furnace design for another hundred years, although the environment may not have been altogether conducive for further innovation. Problems persisted however. Corruption undermined the financial integrity of the business, and success attracted the attention of other rulers wanting the secrets to build their own factories, and unscrupulous individuals who saw the opportunity to gain for themselves at others expense. The stress, the toxicity of some of the chemicals used in his experiments, and the effects of prolonged exposure to heat and bright lights from furnaces took their toll. Böttger died on 13 March 1719. He was found to be deeply in debt through borrowing money on his own account to sustain the factory, but had no descendants and no significant assets (Gleeson 1998, 117).

Darby also needed to juggle his activities, albeit in a somewhat different environment. He too needed until 1713 to achieve commercial viability. Along the way some of his partners sought to withdraw their capital, and extant indentures show that a sum of £1282 was raised from a Bristol financier and paid back between 21 September 1711 and 28 February 1712. In 1714 he was able to renegotiate the lease on the works, for another 21 years from 1717. Two new partners replaced his former ones.

Darby started to expand his operations. Two new operations were established, though probably in partnership with existing furnace operators, in Wales and Cheshire. A new furnace was built on the Coalbrookdale site, and he also recruited three clerks, one to manage each site (Thomas 1999, 17).

Thomas Goldney of Bristol met the capital requirement this entailed through a loan, against a mortgage on the works.

Unfortunately, Darby died in 1717. Coming relatively unexpectedly, and followed by the death of his wife the following year, it did create problems for his family. Nevertheless, the business was divided into sixteen shares, and after some financial manoeuvrings, Thomas Goldney ended up with ten shares and his son one, because of their heavy financial involvement. Three shares were held for the Darby children, and two by the clerk of the Coalbrookdale works, Richard Ford, who later married Darby's oldest daughter. Darby's son regained control in the fullness of time, but in the meantime Ford and Goldney managed the works. In the following decades a very successful business developed.

It must be one of the great ironies of human history that had Böttger found out how to smelt iron with coke, and cast it into pots cost effectively, the King of Saxony would almost certainly have dismissed the idea. For him porcelain was a substitute for gold, and offered a way to further his political ambitions through ostentatious display and military conquest, with the latter being just one avenue for the former. Yet Darby's focus on serving everyday needs, much as with Arkwright and the cotton producers later in the century, and the agricultural improvers contemporaneously bringing about what has been termed the agricultural revolution, underlay the sustainable economic growth that is first observed at this time, and gave the country where it occurred unprecedented political power, until it was overtaken by another that made itself even more adept at catering to people's everyday needs.

The historical evidence of the progress of these two parallel innovative economic developments, demonstrates clearly the impact of the different governance environments. Although major innovations were produced in both environments, both the innovation itself and its long-term economic consequences were shaped by the governance mechanisms available. In both

cases access to investment capital was important, as were commitment mechanisms. Effective legal contracting mechanisms and access to capital played a key part in Darby's innovative activity, while Böttger could only call on patronage and Royal favour. Commitment in Böttger's case was of rather a different nature to that between Darby and his partners and investors. And ultimately it was Darby's incentive to produce for a mass market in a self-sustaining way, over Böttger's incentive to produce for an elite market in a dependant way, which explains why the industrial revolution happened in England, not in Germany. So a hypothesis from our model has survived another test.

The next logical step is to ask whether contracting and finance had changed significantly shortly before Darby founded his company, as if this was not the case there would likely be some other factor recently come into play at that time. It turns out that contracting and finance had both changed, and the next two sections will describe those changes.

3.1 Commitment

The prediction is that there was in England around 1700 a change in the legal institutions supporting contracting. No revolution has previously been identified here, yet when the pieces are put together and looked at in terms of their economic effects, rather than from the perspective of the development of specific legal practices that legal historians take, a remarkable shift did occur at the beginning of the eighteenth century. However, to appreciate what this was it is necessary to trace previous developments from several centuries before.

3.1.1 Legal Environment

By the end of the medieval period England had a complex legal system that incorporated a number of courts, which were to some extent complementary, but also competed with one another. This competition was a considerable spur to innovation, as the different courts competed with one another to

provide legal adjudication, particularly of lucrative commercial disputes, in a way both parties to a dispute could feel was fair and efficient.

English jurisprudence has evolved continuously with no significant dislocations since the beginning of recorded time. Its earliest manifestation was as an integral element of the hierarchical system of administration that became the feudal system. Often labelled the common law system, technically the common law was only one element of the system, albeit the foundational one, that became a fundamental and in many ways a distinguishing element of the English state.

The introduction by Henry II of a common 'traditional' law in the second half of the twelfth century provided a legal environment that was available to all, to some extent. Not surprisingly, the resulting workload was far more than the king could deal with personally, as had previously been the case with disputes that came within his jurisdiction. Given the fees legal action required, this may have been a deliberate revenue raising measure.

Centralisation was not as inconvenient as may at first be imagined, at least from a geographical perspective. This was because the King and his court were by no means a fixed location, but would be continuously travelling around the country. The multiplication of court business, which included local administration, alongside the more limited conception of modern judicial business, was met with the appointment of justices to tour the whole of England, in place of the King. Two were appointed in 1166, but then in 1176 six circuits were established, and by the 1180s twenty or thirty justices were assigned to these circuits. Alongside this, however, there were complaints of the difficulties inherent in searching out travelling courts. So there is evidence that in the late 1170s Henry II ordered five judges to remain at Westminster, where an embryonic Exchequer, controlling the King's treasure and revenue service, had started to put down roots since earlier in the century (Baker 1990, 19, 21).

Alongside the Exchequer, the other great department of state was the Chancery. This had its origins in the Anglo-Saxon scriptorium, where royal charters and writs were drawn up and sealed. The Lord Chancellor was always a senior political appointment, yet most have been lawyers, and in medieval times most were bishops or archbishops who had graduated in civil or cannon law (Baker 1990, 114). However, the bedrock of the common law system was not the chancellors' learning but the documents through which the state was governed, which were made authoritative through the application of the Great Seal of England, possession of which gave the Lord Chancellor his power.

Royal justice was at first seen as a privilege, and one that plaintiffs were expected to pay for. Access to the system was gained through purchasing a writ from the Chancery. By the beginning of the thirteenth century, there were many common forms of writ, and a standard fee for their purchase. Well into the thirteenth century a new form of writ could be drafted whenever the need arose, and registers containing substantial numbers of precedents were put together. By the middle of the century the uninhibited creation of new writs was seen as a burden, and new forms of writs ceased to be invented without the consent of the king's council (Baker 1990, 66).

King John encouraged judicial proceedings before himself, and in 1209 the Bench at Westminster was discontinued. Plaintiffs had to resort to following the King around the country again. This was one of the grievances addressed by Magna Carta in 1215, which in clause 17 set out that 'common pleas should not follow the king but should be held in some certain place' (Baker 1990, 22). The effect of this was that writs had to name the place where the court would be held, and in general this became Westminster, though from time to time it was set for other places. Common pleas were defined as those that did not involve the king. But many issues did involve a royal interest, so in the reign of Henry III, in 1234, the *coram rege* or Kings Bench was revived,

and from this time there were two distinct common law courts, each with their own judges and administrative support.

It is the law developed from the precedents set through decisions in these two courts, Common Pleas and Kings Bench, which technically form the common law. They remained nominally within the purview of the Lord Chancellor however, and it remained possible to appeal to him on issues where no common law remedy existed, or if the available remedies were believed to be inappropriate. To deal with this a court of Chancery developed, and the law applied here became known as Equity. Although initially this spurned the precedent system, and every case was supposed to be decided on its merits, in time the principles it applied became standardised, though never to the extent of the common law. The downside was that proceedings were far more time consuming, and by the end of the eighteenth century a backlog of thousands of cases had built up, and cases could take decades to resolve.

The earliest justices were as much administrators as they were lawyers. Many were clergymen, but some were knights. By the beginning of the thirteenth century the beginnings of a profession was emerging, with new justices being appointed from among those who had previously served as clerks to justices. As the century progressed, standard writs were proliferating, and a body of professionals developed. Oral proceedings were conducted in court French, and formal written documents were in Latin. So, as well as a good understanding of the writs available, effectiveness in pursuing or defending causes effectively in court required considerable linguistic skill.

By the end of the thirteenth century it is not surprising that judges in the two common law courts were appointed from among the professional advocates. During the fourteenth century the system became institutionalised and a guild of serjeants at law was formed, who were recognised as masters of their profession. They came to acquire privileges such as a monopoly on right of

audience at the Common Pleas court, and membership a necessary prerequisite for selection as a judge (Baker 1990, 180).

The rise of a legal profession skilled in the complex reasoning necessary for dealing with the procedural aspects of the law, not surprisingly, came to approach the substantial aspects of the law in the same way. The distinction between procedure and substance was far from clear for many centuries. It was in the course of the thirteenth, fourteenth and fifteenth centuries that the common law was formed from the daily interaction in court between logical reasoning and the real world of plaintiffs and defendants wishing to settle practical disputes. The means by which this was carried out was the interaction between advocates and judges known as pleading. The writ provided the basis on which the case commenced. However, while it set out the basis for the complaint, it provided little which would allow the case to be settled.

A distinction developed between law and facts. This became particularly important because of another development starting at this time, namely the rise of the jury. Early medieval justice depended on proof through supernatural ordeals or oath ceremonies. The end of this process was hastened when in 1215 the church declined to participate in ordeal by fire or water, and their integral role in the ritual ordeal process effectively finished off these forms of proof. The jury system that came to replace this form of proof had its origins in ascertaining the facts in administrative matters, such as the basis on which tax assessments were made. At the end of the twelfth century the practice was still for the judge to examine the jury of local men under oath as to the facts of the case, and on this basis would make his decision.

The rise of the legal profession introduced the practice of extensive debate between the issuing of the writ and the ascertaining of the facts through the jury. The question that the advocates addressed was what exactly the question was that should be put to the jury. It was in doing this that the common law

evolved. A sophisticated process developed whereby the plaintiff put his case, and the defendant then had four choices. He could deny all the facts in a 'general traverse' resulting in the 'general issue.' He could deny one material fact in a 'special traverse' and put a single fact at issue. He could admit the fact but deny that this amounted in law to a case against him, this being a 'demurrer'. Finally, he could admit the facts but explain them through other facts, a 'confession and avoidance.' The first two choices would lead to a jury being called to decide either the general or single fact. The third produced an issue at law, on which the judge could rule. The fourth would effectively reverse the roles of plaintiff and defendant, and the plaintiff had the same four choices in responding to the new factual claims made by the defendant. This switching of roles sometimes occurred a number of times until the issue was decided, which would either be put to a jury or decided in law. In the process, the jury evolved from being local men who would know the facts through their membership of the local community, to an independent body charged with reaching a common conclusion as to the facts of a matter (Baker 1990, 92).

While this was the formal process, the evidence suggests that in practice it was slightly different. Judges were clearly aware that because of the precedent system any formal decisions they made would become law, and they were reluctant to do this, particularly where there was a divided opinion on the bench. Pleas only became binding when the clerks recorded them formally in Latin. The practice developed of making tentative pleas for discussion between the advocates and the judges. So in practice few formal pleas of demurrer were made. Therefore, a plea would be made tentatively, and if the defendant advanced a demurrer that seemed effective, the plea would be withdrawn and another advanced. If no effective demurrer could be found the issue would return to facts, either denied, so initiating the move to the jury stage, or in avoidance, in which case the same process would occur, until either the case was dropped or an issue was reached for a jury to decide.

Given the central role of tentative pleading, the formal records made by the clerks of the court were of little interest, though they were referred to from time to time. It was the actual debate around the tentative pleas that the advocates, particularly when they were still in training, wanted to know about. This need was met through the development of what are now known as yearbooks. Some records survive from the thirteenth century, but it is from 1300 that a continuous stream of reports of arguments in Common Pleas exists in a common form (Baker 1990, 205). They continued in this form until the advent of printing. Clearly, until this time copying was by hand and many lawyers would assemble their own collections. Individual authors only become known at all in the fifteenth century, presumably also a consequence of the introduction of printing. The yearbook collections finish in 1535, but law reports continued, with known authors whose names became their principal identifier.

As well as the change in style of reporting, the sixteenth century also saw a major change in the trial format. Quite why this change took place is not clear, one possibility is that it occurred as an unexpected consequence of an Act of Parliament of 1540. The most significant change was that tentative pleas were no longer allowed, and definitive pleas had to be submitted in writing. The practical effect of this was to reverse the order in which issues of law and issues of fact were decided. Two specific changes occurred to enable this reverse to take place. First, judges needed to be willing to be authoritative and to rule definitively in the case of demurrers. The second was for a formal appeals process to develop. In the late fifteenth century it became possible to raise motions regarding substantive questions of law after the verdict, and over the next two centuries this process went through further elaboration. This format has remained in existence ever since.

Given this rather extensive explanation of developments in the common law legal process in general, albeit one necessary to understand the succeeding,

specific developments in what would become known as contract law will now be considered.

3.1.2 Rise of *Assumpsit*

The genesis of modern English contract law is in the action of *assumpsit*, or undertaking, which was treated as an action with its own identity from the sixteenth century. It differed from the ancient actions of debt, detinue and covenant because the plaintiff claimed damages, rather than recovery of the debt, chattel or charter (Simpson 1975, 199, 210). Because the appreciation grew that it did not require the existence of a previous act, it became possible to take action on a promise, that is, an agreement between two parties, while the ancient actions could only require the agreed reciprocation or reversal of an already completed act.

The history of *assumpsit* can be traced back to the fourteenth century. It was just one way of pleading a trespass action for damages because of a wrong, or in modern legal terminology a tort. Writs of trespass were designed to compensate for actions that could no longer be put right, so were irreversible (Simpson 1975, 200). Around 1370 lawyers began to draw the distinction between general trespass and 'writs of the case.' The former came to relate to a breach of the peace, the second had two elements, the first 'the case' that specified a duty the defendant was expected to perform, and the second element then alleged the way in which they had failed to discharge the duty (Simpson 1975, 202). Prior to this time it had been standard practice to allege breach of the peace regardless of the actual circumstances, because this was what qualified the action to be heard at the royal courts, rather than the local ones, because the king had a specific interest in keeping the peace. However, towards the end of the fourteenth century the courts came to regard frivolous allegations of breach of the peace as objectionable (Simpson 1975, 203).

The duties that defendants of trespass actions were alleged to have failed to perform at this time arose from local tradition, such as maintaining a sea wall,

or from an individual's status as a member of a particular trade, such as a doctor or a blacksmith. For example, before the latter part of the fourteenth century a plaintiff who alleged that a blacksmith who incompetently put a nail into a horse's hoof, the wound becoming infected and the horse dying, would simply claim that the blacksmith killed the horse. However, after this time the court would expect an acknowledgment that the plaintiff had agreed with the blacksmith for his horse to be re-shod. This though opened up the opportunity for the defence to claim that this should not be brought as a trespass, but as a breach of covenant, which would not allow damages to be claimed, and only recognise that the agreement had not been carried out properly, the remedies for which were that it be unwound or completed, the monetary value of either of which would be negligible. So plaintiffs started to play down the element of prior agreement. The courts however also started to accept that the issue should be the wrong committed, not whether there was a prior covenant or not. However, to avoid using the term covenant, the term *assumpsit* was first recorded as being used in 1373, and came to be common practice soon afterwards (Simpson 1975, 215).

Through the fifteenth century, *assumpsit* was used in the context of agreements, but the distinction was made between simple failure to carry out an agreement, which still required action on a covenant, and causing loss to the plaintiff through negligence. This was known as the nonfeasance rule, which first appeared in 1400 and was still quoted in 1533, albeit in a dissent from the majority opinion (Simpson 1975, 225). In many cases the initial agreement was of a relatively minor nature, so the formalised action of sealing a document that was required for creating a covenant would not be gone through. The fact that work had at least begun to be undertaken was however evidence that the agreement existed (Simpson 1975, 224).

The nonfeasance rule did not disappear overnight. According to Simpson (1975, 264) the first half of the sixteenth century is very poorly covered by law reports, so a definitive assessment of when this happened is impossible.

Nevertheless he is able to show that there was a relatively gradual move away from the rule.

As the sixteenth century progressed, two new problems came to concern lawyers. The first was to bring the action of debt within the purview of assumpsit despite an, until then, well-established distinction between the two. Secondly, in some ways a more genuine problem, was establishing the basis on which the informal agreements that had not previously been a concern of the royal courts should be recognised (Simpson 1975, 271). The records of three cases, from 1527, 1543 and 1544, show that the second issue was becoming recognised, but found no effective way of dealing with it. Nevertheless, assumpsit was becoming an action for breach of promise (Simpson 1975, 274).

The availability of assumpsit as an action for breach of promise was in part due to the lack of available remedy in other parts of the common law. In particular, for specific types of contract, particularly in cases of sale of land or for performance of a service, where one party could sue for debt, but the other had no remedy. It was also becoming more widely used as a consequence of forms of pleading. Debt's roots in traditional society meant that the way of rebutting a plaintiff was through 'wager of law,' and juries were still not utilised in these cases. This meant that to deny the charge the defendant simply assembled twelve men each prepared to swear that the allegation was unfounded. If each one took the oath, with no slips of the tongue, then the defendant was absolved. In assumpsit a jury took the decision as to facts. Again change did not occur overnight, but by 1602, with Slade's Case, assumpsit had won out.

The shift from debt to assumpsit was further complicated because of the two common law courts where plaintiffs could choose to pursue their case, the Kings Bench and the Common Pleas. In theory Common Pleas had a monopoly on personal actions, while the Kings Bench required some royal

interest, although for these types of cases this could be achieved when the defendant was in the court's Marshalsea Prison, or if the case involved one of its officers. However, it became standard practice simply to allege that the defendant was in the custody of the marshal, from which point Kings Bench would hear the case. This became significant because it was generally regarded that the procedure in Kings Bench was better, where parties could be represented by utter-barristers, while in Common Pleas expensive and often elderly serjeants-at-law held the monopoly of right of audience. This trend is confirmed by evidence that towards the end of the sixteenth century Common Pleas business was depressed (Simpson 1975, 292). Two cases in Kings Bench show the evolution from debt to assumpsit in progress. In 1533 a plaintiff claimed that he had suffered greater damages than simply the value of goods not delivered, because in order to stay in business he had needed to pay more at short notice to another supplier. The jury found against the plaintiff on the facts, but the plaintiff raised the issue of law that he should be using debt. The Kings Bench, in upholding the jury's decision, denied that liability for debt excluded the possibility of liability for assumpsit. Then, in 1542 a judgement was given that a judgement in debt bars assumpsit, and vice versa, so recognising that they were alternative cases on the same cause (Simpson 1975, 290-1).

One of the reforms that helped bring oral pleading to an end, a Statute of 1585, established a judicial appeals mechanism from the Kings Bench, to replace appeal to Parliament. The way the appeals tribunal was constituted meant that in effect judges from the Court of Common Pleas dominated it. The Kings Bench however did not recognise the appellate body as authoritative. This created the situation that for a number of years at the end of the sixteenth century Kings Bench would routinely allow actions of assumpsit on informal contracts, which the appeals tribunal would then disallow (Simpson 1975, 293). The traditional remedy of a conference of all senior judges was invoked to address this situation. Slade's case, originally from 1597, was the specific case on which the issue was debated. The

resolution, although never stated formally, was that subsequently judges accepted assumpsit in debt actions, as an alternative to defending a debt action through wager of law. Chief Justice Holt was still upholding this basis for the distinction in 1689 (Simpson 1975, 310). Slade's Case thus effectively abolished wager of law as an institution, although it was not until 1833 that a statute formally abolished it (Simpson 1975, 298).

So alongside the issue of moving from debt to assumpsit for action against breaches of informal contracts, the other issue was the basis on which the agreement should be recognised. What should replace wager of law? The answer was consideration. The basis of an enforceable promise became that the promisor had received some benefit that prompted them to make the promise. The desire for a simple dictum was satisfied by this apparently simple test. Of course, the varied and complicated reasoning that may be behind a promise ensured that the test was never as simple as it at first appears.

Not surprisingly, the doctrine of consideration appeared at the time assumpsit became available as an alternative to the action of debt, around the mid sixteenth century. In the fifteen fifties consideration started to be averred, 1549 actually being the very first, but at this stage it was by no means inevitable. Then in a series of cases in 1566, 1567 and 1568 the actionability of agreements was reported as being clearly related to the presence or absence of consideration. Subsequent reports show the development of an intricate doctrine and by 1586 a Solicitor General identified the three key issues as consideration, promise and breach of promise. Simpson (1975, 319) believes that by the end of the sixteenth century the major problems relevant to the issue of consideration down to modern contract law had been raised and settled, one way or the other. Nevertheless, on his own analysis, while the central importance of consideration had been established by that time, issues such as precisely who was able to sue, and that one promise was sufficient consideration for another, had to wait until well into the seventeenth century

to be settled, the latter using rules set out in a case of 1669, the former using reasoning from a 1677 case (Simpson 1975, 482, 465).

Also in 1677 occurred what Simpson considers to be the most significant development in the common law of contract until at least the mid nineteenth century, if not today, namely the Statute of Frauds (Simpson 1975, 620). In some ways this was a revisionist piece of legislation, because it reintroduced the need for formalities in a number of areas that the development of assumpsit doctrine had allowed to be pleaded with fewer restrictions. However, the effect was to shift the balance from the plaintiff back to the defendant. Contemporary rules of evidence meant that often one side simply pleaded against another, sometimes with the evidence of witnesses who were not cross-examined. It was generally agreed to be common for people to sue for breach of promise on the least pretext, with success often coming down simply to a matter of luck. The Statute was not unduly onerous; it simply required that for promises in a number of cases the evidence be in writing, including contracts with a value of over £10 in value or one year in length.

The nature of the common law meant that the medieval actions did not go away. Parts were superseded, but other parts assimilated, and some left untouched. Despite assumpsit's rise to prominence in the sixteenth century and its status as the regular common law remedy for broken agreements, it was not until the middle of the eighteenth century that the extant evidence of legal practice ceases to be made up primarily of material on the medieval law (Simpson 1975, 3). The following quarter century saw the contract law developed up until that time firmly entrenched, through the work of Chief Justice Lord Mansfield, who brought what had been issues of fact to be issues of law, so that rather than rely on the somewhat unpredictable nature of juries many issues became clear and predictable, as he believed was wanted by the business community (Simpson 1975, 618).

So by the middle of the seventeenth century, through a process of gradual evolution, a sophisticated system of rational contract law had developed. A clear distinction was established between issues of law and issues of fact. Issues of law could be expected to be well known to businessmen and provide a high level of predictability in the system, and while formality was encouraged, it was still possible to provide remedies in the event that the agreement had not seemed to warrant the cost of formality, or the issue causing the problem had not been foreseen. The downside was that the rate of evolution was, in practice, painfully slow. The next two sections describe areas where this became a particular problem, before describing how the solution came about.

What is apparent from this and the previous section is that the procedural and contractual logic described here is in most cases far from obvious. Some medieval innovations such as wager of law are obviously problematic from the modern standpoint. Others such as the doctrine of consideration remain largely because they work. Even issues such as the separation between law and fact, while clear in principle, are not always so clear in practice. It is the economic consequences of many of these practices that are their ultimate justification, yet this is not a well-researched area, for either lawyers or economists. However, before considering the financial and economic consequences of these developments, it is worthwhile to consider further legal developments.

3.1.3 Legal Organisation

The legal form of business organisation had two sources. At the end of the medieval period, partnerships were the only ongoing contractual arrangement that could be formed independently of the state, and they owed their origins to classical developments, largely described above. Other forms of organisation had medieval origins, but not as businesses, rather as ecclesiastical bodies or guilds. The latter, although created to support craftsmen in business, acted in regulatory, welfare and social roles rather than

businesses in their own right. The beginning of the sixteenth century saw the creation of the regulated company which was in some ways similar to the guilds in being an organisation within which merchants traded on their own account, but they were limited to providing business support services such as foreign bases, as an aid to moneymaking, and had no significant welfare or ritualistic elements like the guilds (Harris 2000, 32).

The first joint-stock company, the Muscovy or Russia Company, was founded in 1553, and chartered in 1555. Soon afterwards other companies formed for long distance trade with other specific destinations. Although first organised with permanent capital, the Russia Company did not reach profitability, and further calls were made on shareholders, to the extent that in 1568 it was reorganised using one to three year short-term accounts rather than perpetual capital. Then in 1622-3 separate term accounts were replaced by individual accounts, and it became a regulated company. The time was clearly not yet right for the joint-stock company (Harris 2000, 45).

The first companies were created with a symbiotic relationship between themselves and the monarchs who endorsed their charters, which involved the exchange of expected company revenues for monopoly rights (Harris 2000, 50). Although from a modern perspective monopoly rights could be seen as a mechanism to ensure that initial investments in establishing relationships and facilities in foreign locations would be protected from competition from subsequent free-riders, they were seen then simply as a way of protecting profits, and were possible because the royal prerogative would provide effective protection from competition. The monarch was happy to intervene in this way to generate revenues for himself, particularly during the accelerating conflict between king and parliament during the first half of the seventeenth century. This relationship between state and company survived the 1688 revolution and, as the next section describes, the joint-stock companies became major players in the management of public finances.

The post revolutionary period saw a burst of company formation, something like 150 in 1695, for a whole range of purposes, including manufacturing, mining, banking and finance, fishing and water supply. However, although total capitalisation continued to increase, actual numbers decreased to around 46 in 1693, and only twelve in 1717, with only four new companies formed between 1700 and 1717 (Harris 2000, 57). Of an estimated £20.6m in share capital in English companies in 1717, £18.7m was in the Bank of England, the East India Company and the South Sea Company. It is clear that for the three, government debt finance was their primary business, so it is likely the same held for the remaining capital. So, while the market in company shares looked attractive, and there was considerable activity, it was directed at public rather than private investment. So it appears from this that company forming was not a key to growth. But growth did lead to a desire to be able to create appropriate legal structures, which led to much experiment later in the eighteenth century, described below in the context of implications rather than causes of sustainable economic growth.

3.1.4 Bills and Notes

Bills of exchange and promissory notes were of fundamental importance in the eighteenth century English economy, as the next section demonstrates. It is instructive to build on the descriptions above of the common law legal environment and the rise of the action of *assumpsit*, and observe the development of the legal framework that facilitated this use of bills and notes. The precise pleas submitted to the courts plot out in great detail the path developments followed.

From the end of the sixteenth century, payees of bills were succeeding in *assumpsit* actions against the drawees with an explicit allegation that a sum of money had been delivered as part of the exchange transaction of which the bill was a part (Rogers 1995, 135). Soon afterwards pleas start to cite the custom of merchants, which is not straightforward to explain, because lawyers would not advise their clients to claim more than necessary, as the more that

is claimed, the more there is for the defence to challenge. Rogers (1995, 133) speculates that this may be because it is not clear which party is the agent and which the principal, so if in earlier cases the agent had contracted on behalf of the principal, the principal was bound to pay. However, the principal could not bind his agent in the same way, but it could be that by asserting that it was the custom of merchants to act on earlier promises, particularly a promise to accept the bill for later payment, the principal's promise could be binding on the agent.

Developments in the seventeenth century can be seen in both official records of pleadings and reports of exchanges in court. A report of 1612, although brief, sets out the custom of merchants, but does not mention any underlying exchange. Then in 1636 there is a record of a formal plea that takes this approach. The 1690s have several records of pleading where the practice is simply asserted as being according to the custom of merchants, with no detailed description of what this is (Rogers 1995, 129-131). Rogers notes that this continuing reference to the custom of merchants has been taken to mean that this was alleged as a statement of fact. From time to time defendants demurred on the basis that they were not merchants, but the courts rejected these pleadings as a matter of course. The explanation Rogers proposes is that it was a typical common law device that allowed the law to develop to meet the needs of the economy. Technically, the promise to pay would require either a sealed document or a consideration, but neither of these existed. The assertion of custom of merchants provided a way to enforce what clearly had become the general expectation in this particular type of economic transaction (Rogers 1995, 149). The common law was continuing to respond, albeit in a slow and often convoluted way.

3.1.5 Parliamentary Innovations

From a modern perspective it seems curious that in all this discussion of the legal environment, and of the law as it relates to contracts in general, and financial instruments in particular, there has been so little reference to

Parliament and statute law. Of references so far, two sixteenth century Statutes merited mention as a result of apparently unexpected changes they brought about, and the 1677 Statute of Frauds was on the face of it a retrogressive Act. So even where mention has been made it is not in a particularly constructive role. Nevertheless, it will be useful to consider briefly the background and role of parliament at the beginning of the eighteenth century.

The courts and parliament both had their origins in the same royal council that advised the king. In medieval times it was settled that the canon of statutes began with Magna Carta on its confirmation in 1225. For the next century a variety of formulas are used to introduce statutes. Before the reign of Edward III (1327-1377) parliament and council are interchangeable terms. Reference to lords and commons had been made since 1275, but during the reign of Edward III the commons began to meet as a separate house, and in 1340 the king conceded the principle that without their consent no taxation could be imposed. From 1348 it was said that the king with the assent of the lords and commons made laws, and from 1407 until today it has been accepted that English parliamentary legislation requires the separate consent of the monarch, lords and commons. Official records of legislation only date from the 1290s, and until the fifteenth century officials sometimes carried out the detailed drafting after assent had been given. It was only at the beginning of the seventeenth century that the text of the statute became the overwhelming issue when disputes came to court (Baker 1990, 235).

For subsequent developments to be understood, what has become the third element of modern government, namely the executive, needs to be considered. Again from a modern perspective, at least as important as the role of the law in resolving disputes between individuals, is its role in mediating between the individual and the state in the form of the executive. However, the early medieval origins of both the judges who developed the common law, and of parliament, were in the king's council, and the king was in no

sense bound by his council. It took a civil war and a bloodless revolution during the seventeenth century to establish the rule of law over the king. Until the Bill of Rights of 1689 established the limitations on the prerogatives of the king, many seventeenth century developments took place through the interaction between royal prerogatives and parliamentary powers.

The most obvious area of royal prerogative relative to the issues considered here is that of monopoly. The common law drew from the right in medieval custom of everyone to ply their trade, that no one could claim a legal monopoly to prevent others from setting up in competition. However, royal charters for markets and fairs, as well as for guilds and trade associations, effectively created monopolies for their holders, but the purpose of these was to set standards and provide an environment in which any qualified individual could ply their trade. However, in the sixteenth century the practice was adopted from the continent of granting monopolies to individuals who introduced new inventions into the realm, the argument being that this infringed no existing trades. This was extended to grants of monopoly on trade in foreign parts, by virtue of the royal prerogative to license foreign trade, so again infringing no existing trades. It was this that allowed the trading companies to be chartered.

Grants of monopoly became a lucrative source of revenue for the crown, and the boundary was continually extended. The issue came to a head in 1602 with the grant of a monopoly on supplying playing cards, the crown arguing that it was an undesirable activity requiring control. The courts dismissed this defence however. Nevertheless, the practice continued and action was taken in Parliament leading to an Act of 1624 declaring any monopolies within the realm void, with a number of exceptions. Most long lasting were rights of inventors, who could claim, as Darby did, a monopoly for fourteen years. Charters to companies were also excluded for the time being, so giving an extension to the monopoly rights of the trading companies (Baker 1990, 513).

The interregnum following the execution of Charles I saw a vast range of sweeping legislative changes, but these were all rolled back on the return of Charles II in 1660. However, one change that was seen to have benefits was the end of feudal tenure, so this was enacted in a Military Tenures Abolition Act of 1660 (Baker 1990, 294). The Statute of Frauds of 1677 has already been considered, but was another intervention at this time. Following the revolution of 1688-9, mineral rights, with the exception of gold and silver, no longer belonged to the crown, to the undoubted benefit of the landed classes, but also removing administrative barriers to the exploitation of these resources about to become a central feature of the industrial revolution (Atiyah 1979, 87). Parliament was beginning to take the legislative initiative, albeit in relatively unproblematic areas from an economic point of view.

The start of the eighteenth century sees a change of gear. In 1702 Chief Justice Holt ruled that a promissory note payable on order was not a bill of exchange, and hence not negotiable. His view was that it would in effect take the place of the sealed covenant, and any written agreement would then become binding in itself, with no dispute possible on the agreement of which it was a part (Baker 1990, 418). Three subsequent cases upheld this ruling. Rather than wait for the common law to take its course, and find a way over decades to deal with what from an economic point of view was a useful new tool, as had happened previously, in 1704 the Promissory Notes Act made all notes payable to "X", or to "X or order," or to "X or bearer" negotiable (Carruthers 1996, 130).

3.1.6 Legal Revolution

While not perhaps as direct a consequence of the Glorious Revolution as was the financial revolution, to be considered shortly, the final acceptance of Parliament, or technically the King in Parliament, as the sovereign body of the nation in many ways brought about the end of the common law tradition of the law adjusting internally, albeit to the changing demands of the society it served. From the beginning of the eighteenth century, a body that had seen

itself finally put an end to unrestricted royal prerogative had the confidence to legislate when conditions warranted it. Clearly, the common law system had not been so explicitly problematic as the previous executive system and its financial manipulations. But nor did it have continental examples from which it could easily copy. For these reasons the effects were not as apparent as in the financial arena, but they were non the less significant.

Through the sixteenth and seventeenth centuries a necessary shift in people's, and particularly the law's, understanding of contract had taken place. The established idea was that the reason for the contract was critical, whether based on an existing relationship or transaction, and the law made the contract. The new concept was that a contract was an act of will and was created by the individuals involved. By the mid eighteenth century the realisation grew that there was a distinction between morality and the law of the land, that the two were not necessarily the same (Atiyah 1979, 163). As people came to rely more on one another as individuals in their own right, rather than as holders of a place in society, they came to expect that they should be able to make commitments to which they would be subsequently held liable, and the law developed to make this possible (Atiyah 1979, 188).

3.2 Credit

The fiscal effects of the political changes of 1689 have already been labelled as a financial revolution (Dickson 1967), referring principally to the getting and spending of money by the state. Here this is labelled 'public finance', although clearly taxes were collected or money borrowed from individual people, and given to individual people in return for goods and services. But, in addition to this, I want to argue that this revolution was larger in scope than conventionally envisaged, and also took place on the level of private finance. The level labelled here 'private finance' concerns the transfer of purchasing power between individuals, although again changes here also involved changes in national policy regarding legislation and judicial practice.

Furthermore, as in the commitment-credit model used to derive the hypothesis being tested here, contracting mechanisms and financial markets are self-reinforcing, and most of their economic effects come from their interactions. So the distinctions made here, dividing credit and commitment into separate sections, are to ease the process of exposition, more than any desire to claim that separate but parallel developments occurred.

3.2.1 Public Finances

Although public finance did not become truly self-sustaining until the twentieth century, when it shifted away from mainly paying for unproductive national defence to mainly investing in human capital through education, health and social security, seventeenth century developments, first in the Netherlands and then particularly in England, did enable it to become effectively self-sustaining. Governments have always needed to borrow to finance calls on expenditure that cannot be met immediately by taxation and other forms of revenue generation. What was new at the start of the eighteenth century was the expectation that the debt would be funded, that is repaid, so leading to a low risk premium and low interest rates, which made it affordable by governments. This created a self-reinforcing system, because reliable repayments kept interest rates low, and low rates meant governments could afford to pay, and had the incentive to do so to keep rates low.

However, what really allowed the system to be sustainable, though not recognised at the time, was that the public debt markets that developed provided liquidity and a 'risk free rate of return.' This provided a pool of funds and a standard by which to compare other lending opportunities, and allowed businesses to borrow. Clearly, this would be at a risk-adjusted rate but one not much higher than the rate for government debt, at least for well-collateralised businesses. This created a very conducive climate for commercial investment, and brought about economic growth. This growth in turn increased the economic surplus available to the government, and even

with constant tax rates this meant increasing revenue, thus funding the previously incurred debt in a sustainable way.

Central to this revolution in public finances was the constitutional settlement following the Glorious Revolution of 1688, which invented the idea of parliamentary sovereignty. Public finance was no longer the battle, at times literal, that it had been over the previous century between a royal sovereign demanding spending and a parliament trying to resist the increase in taxation this implied. In this struggle, royal borrowing was continually filling the resulting fiscal gap. But the frequent defaults that followed made government borrowers one of the most risky. Normally this was compensated for through the granting of a monopoly, or similar wealth destroying profit opportunity, further damaging the sustainability of the system. 1672 saw the last major sovereign default in England, when Charles II's pursuit of wars with the Dutch proved economically unsustainable.

Following the 1689 settlement, the new joint sovereigns, the Dutch William and his English wife Mary, daughter of Charles' brother and successor, the now exiled James, still led the executive; although in the decades and generations that followed this became an increasingly symbolic status. The executive itself came to be made up of members of the party that had the majority position in the House of Commons. Spending had to be initiated by the executive, but approved by Parliament, which also carried out retrospective audits (North and Weingast 1989, 816).

North and Weingast (1989, 821) pick out four specific practices introduced at this time that were central to the public financial revolution. First, all new long term loans had specific taxes earmarked for paying their interest. Second, it was proposed that a Bank of England be incorporated. Although the first large long-term loan secured on specific tax revenue, raised in 1693 to pay for a war with France, was fully subscribed, by the next year it was thought that sources were largely exhausted. So, following the Dutch example, it was

proposed the new Bank would have a dual purpose. First, it was responsible for disbursing funds to government suppliers, and second, subscriptions would be sought which would be lent to the government. The dual role meant that should the government ever default on an interest or capital repayment, made in the first instance to the Bank and only subsequently to its shareholders, the Bank would immediately cease payments to government contractors. This created a strong incentive to the government to live up to its financial commitments, to avoid problems with unhappy suppliers refusing to continue trading with them, or taking steps to recover monies owed.

So successful was the proposed Bank that one third of its initial capital was subscribed within one day, two thirds within the next two, and the whole amount ten days later. Two subsequent innovations further strengthened the financial system, the creation of a separate 'sinking fund' to make up shortfalls in taxes earmarked for a particular loan issue, and milling of coin edges to reduce the practice of shaving coins, which debased the currency.

The effects of the public finance revolution were startling. Previously government debt had been limited to around a million pounds, between two and three percent of Gross National Product (GNP), with government expenditure double that. Afterwards, Britain was able to follow the Dutch example, where long-term government borrowing cost four percent a year, while in England only short term borrowing was possible, and at rates varying from six to thirty percent annually. The initial loans of the 1690s were at fourteen percent, which halved by the end of the decade to between six and eight percent. Within the first ten years, expenditure had quadrupled to £7.9m and debt was at £17m. Remarkably, there was no significant price inflation over this period. By the 1730s interest rates were as low as three percent (North and Weingast 1989, 822-3).

Two consequences followed, one political and one social. Both of which had important economic effects. Politically Britain, as England became in 1707

following the Act of Union with Scotland, became the dominant European and then World power during the eighteenth century. To a large extent this was a result of the extraordinary way the state was able to raise finance to pay for ever more expensive wars. As North and Weingast point out, by 1765 economic conditions may have been superficially similar in Britain and France. But while France had eroded its capital base to the extent that it was on the verge of bankruptcy, Britain had strengthened its capital base to the extent that it was on the verge of the industrial revolution (North and Weingast 1989, 831). Not unrelated to this was the rise in influence of what was called the monied interest, that is, those involved in business, as a political force. This was in contrast to the previous dominance in political society of the landed interest, aristocratic owners of large tracts of farmland. Commercial considerations became significant, and geo-strategic policy was dictated at least as much by the need to protect and further commerce as it was by national security in the more conventional sense of preventing foreign invasion.

The social consequence of note was what could be called the de-personalisation of credit. Prior to the Glorious Revolution, loans to the state were personal loans to the King, with all the additional complex issues that go with personal relationships. Once the Bank of England was established, lending to the government became to a much larger extent limited to the explicit contractual issues associated with the transaction. While the empirical evidence cited demonstrates how big and fast an impact it had on public finances, the impact was, if less easily demonstrable and less immediate, at least as great on the private sphere. As people became used to impersonal public debt, so they became open to private impersonal debt. The rise of private banks to provide the shield of impersonality played a significant role. Government debt provided an interest rate benchmark, and bankers learned to measure commercial lending opportunities against this. So long as the bankers maintained and serviced people's deposits, they did not much care

where they were used, and entrepreneurs with collateral could borrow to invest with no encumbrance other than those explicitly in the contract.

However, as noted above, this was only one, albeit important, side of the story. While political developments had been proceeding rapidly through the seventeenth century, so had commercial practice.

3.2.2 Private Working Capital

In much the same way that, in retrospect, public finance went through a revolution at the end of the seventeenth century, but probably at the time appeared more like a gradual evolution than a sudden revolution, the same pattern occurred in the private sphere. It will help to start by drawing a distinction between working capital and investment capital. Working capital within the economy is to a large extent the same thing as the money supply, which expanded hugely in this period. It enables transactions to take place, by having some form of money or debt given in return for goods or services, which in turn can be used to purchase other goods or services. Investment capital markets, where purchasing power is temporarily transferred to purchase or develop assets that play an intermediary role in producing other goods or services, took a little longer to develop, and will be considered in the next sub-section.

In many ways the foundations of developments in England were laid in the middle ages, and the path from the Italians through the Dutch to England was sketched out in the previous chapter. What was new in England was de-personalisation, and particularly the strengthening of negotiability of financial instruments that essentially were able to become a form of currency. There has been a considerable amount of confusion over the process by which this occurred. Rogers suggests that because negotiability of instruments, that is their creating contractual obligations in their own right apart from the transactions of which they form a part, is central to modern legal practice, it was assumed that this was so in the past (Rogers 1995, 2). The confusion was

helped because many earlier writers, particularly in the eighteenth century, were making assertions regarding credit practices in support of essentially political arguments, but these assertions were there to add weight to their argument, rather than provide a dispassionate account of the relevant legal principles (Rogers 1995, 155, 157).

By examining case law records, both official records and various digests, Rogers is able to document the changes in the use of bills and notes before and after the middle of the seventeenth century (Rogers 1995, 94). In the prior period the tone utilised in works describing their use is one of a specialized and arcane activity. But by 1687 a Chief Justice even describes himself as being directly familiar with them through his own use of them (Rogers 1995, 96-7). Rogers makes the point that developments in common law arise through responding to the need to adjudicate practical disputes, so changes were to a large extent a result of changing practices of businessmen in the economy. Nevertheless, it is important to note a key shift from the early period. From their earliest use, bills were integral elements in foreign exchange transactions, and served the purpose of facilitating international trade in different currencies and locations. A merchant would sell his goods in a foreign city and give the money to another merchant trading in the opposite direction, in return for a bill payable to the merchant's agent or principal in the merchant's home city, by the agent or principal of the merchant to whom he has supplied the hard currency, so transferring hard currency in the other country in the other direction, from the person who sold goods there, to the one who purchased them. The benefits from being able to do this were considerably enhanced by the growth of large ports and the existence of brokers who would match up merchants, based on their reputation for knowing the reputation of the merchant to whom hard currency was being supplied.

After the middle of the seventeenth century a significant change took place, with the development of 'inland bills' where both parties were in England.

The earliest examples date from the beginning of the seventeenth century (Rogers 1995, 101). This can be explained by the rise of middlemen, with a shift from a largely local to a national economy. A good example of this is the Blackwell Hall market for woollen cloths in London. Established as early as the late fourteenth century, there was a major change in the seventeenth century. Previously, the clothiers who organised the manufacture of the cloth had dealt directly with the drapers who either exported or sold it to consumers. But further economic specialisation occurred, with the development of specialised product factors, who had clothiers ship them the cloth, which they then sold on a commission basis to the drapers (Rogers 1995, 104).

There were two gaps that needed to be filled for the system to work effectively. In most cases the cloth was manufactured a considerable distance from London, so while the clothiers' expenditure would be in their home region, their income would be in London. Secondly, the drapers would often not be able to pass on the cloth for some time, so there was a considerable time lag from the expenditure on production to consumption or export. The factors filled in these gaps because they would extend credit to the drapers, albeit on behalf of the clothiers, to whom they would then also extend credit. This credit was, however, in London, not in the clothiers home region. The clothiers turned this situation to their advantage. As the major national trading centre, and the centre of government, funds flowed back to London from other areas in two ways, through tax remittances, and through local merchants who purchased goods in London, often from abroad, and sold them in their home region. What the clothiers were therefore able to do was to exchange the hard currency collected by tax collectors and local merchants for bills that drew funds from their account with their factor in London. The tax collectors and merchants could then avoid the costs and risks of transporting cash to London. They simply presented the bills given them by the clothiers to their factors, who gave them the cash to pay into the treasury or to their own suppliers.

The inland bill system really took off when it became the practice accepted by the courts that bills should be honoured regardless of any complications arising from the transaction of which they were a part. In this way they became currency in their own right. The practice grew up that people would use a bill payable to them, or their nominee, or even just the bearer, as currency, where necessary endorsing the bill as payable to the person they were transacting with. It became common for bills to be endorsed many times before finally being cashed in with the person on whom they were originally drawn. Clearly for this to happen there needed to be a good deal of confidence that they would be accepted once they reached their destination. The distinction between a bill and a note arose because a note was payable to the bearer and it was not possible to go back to the person from whom one had received it, while a bill could in theory be unwound by going back up the endorsement trail, though clearly this could not happen very often for the system to work.

Clearly the person who took a bill in exchange for cash or goods or services did not have to do so at its face value, they could discount it, and the difference between the face value and the discounted price would serve as their commission and effectively as interest. In this way many cloth merchants, particularly in Lancashire as early as the 1680s and 1690s, built up complementary businesses, and in some cases regularly sent cash to London hidden with their cloth, because they were selling bills in greater amounts than they were earning simply from the sales of their cloth (Rogers 1995, 114).

At the London end the goldsmiths, many of who were by this time becoming bankers in all but name, spotted another business opportunity. Their expertise had made them successful in the business of exchanging foreign and domestic coins, which had extended into the purchase of foreign bills of exchange. The move to accepting, at a discount, inland bills was the next logical step (Davies 1994, 251). This line of business was complemented by their role as deposit takers. Carruthers shows that at the time of the stop on the exchequer in 1672

the twelve biggest creditors were all goldsmith bankers, who accounted for 97.5% of the total sovereign debt (Carruthers 1996, 63). They were clearly preferred by ordinary investors for investing savings as an alternative to the state, although in many cases the goldsmith bankers simply put the money to use by lending to the sovereign. Having said this, no doubt one of a small number of large debtors was perceived as having more chance of repayment than one of a large number of small debtors, given the personal nature of sovereign debt at this time. With the creation of the Bank of England, competition for deposits opened up (Carruthers 1996, 141), although goldsmiths and private bankers needed to seek competitive advantage through lending to private rather than public clients. The Bank of England also became an active discount house, although this did create somewhat of a conflict of interest that needed careful handling from time to time (North and Weingast 1989, 827). The final step towards modern fiat currency was taken when goldsmith bankers, and shortly after its formation the Bank of England, started to loan money in the form of promissory notes, which paid the bearer metallic currency on demand, but was clearly a more convenient and often a more secure medium of exchange. By 1776 Adam Smith asserts that the level of bank money in circulation clearly exceeds that of metallic currency (Smith 1776, 389).

Banks as autonomous institutions outside London took some time to become established. Davies believes that the evidence is that there were only around twelve by the middle of the eighteenth century, including two in Bristol where Abraham Darby first started in business. But by 1775 there were over 100, and growth was rapid after that. After 1808 banks were required to have a licence to issue notes, and the peak of licences quickly followed in 1810 with 783 (Davies 1994, 286). It is interesting that although Lancashire was one of the first places for inland bills to be widely circulated, it was one of the last places to develop local banks. Even in the late eighteenth century the primary circulating medium of the region was bills, over either Bank of England, or any other bank's, promissory notes (Rogers 1995, 123). An Act of Parliament

in 1708 clarified a previous act, and granted the Bank of England the national joint-stock banking monopoly, in return for taking on much of the government's debt. The 1708 act forbade associations of more than six people from carrying out banking activities, and this remained in force until 1826. So there was no real possibility of consolidation until after this date.

So the public finance revolution was only half the story, although public and private finance were closely interconnected. The working capital that largely oiled the wheels of the industrial revolution was made available, either through discounted bills, or later through local banks, because it could be easily priced, at a level that was economic for manufacturers, as a result of the benchmark provided by government debt.

3.2.3 Private Investment Capital

Even in the mid eighteenth century, calculations suggest that the ratio of fixed to circulating capital in British industry and commerce was 1:1, rising to 3.3:1 by the mid nineteenth century (Daunton 1995, 259). So working capital was certainly more important around the turn of the eighteenth century, which explains why most developments were in working capital rather than investment capital.

Nevertheless, fixed capital was important. To some extent it could be hidden within working capital, and Darby's short term borrowing to support the development of his process appears to use a mechanism more naturally applied to working capital, presumably because that was what was available. His initial reliance on business partners shows where the majority of his investment capital came from. It is possible that partners did not even need to have funds immediately available, as in the case of Guttenberg, although probably unlike Fust in fifteenth century Germany, loans would not have been secured on personal reputation but on the security of existing businesses, and particularly on property. That this was becoming more important is shown by a 1704 Act of Parliament to establish a deed register

for West Yorkshire, copying another Dutch invention that later spread more widely across Britain. Because, as the Romans appreciated, securing loans on property is all very well, but how does a lender know that a potential borrower has not already used their property as security for a loan with someone else? The legal device of selling the property, subject to the right of the vendor to repurchase at a fixed price in the future as well as maintaining use was utilised (Daunton 1995, 244), but is clearly messy and not what is really required. A deed register which is easy and cheap to use, particularly when it has the stick as well that the courts would not enforce a lien unless it was on the register, which is what the 1704 Act did, was a simple and elegant solution to the problem, and was soon copied.

Again, legal innovation provides a reliable and enlightening source of evidence regarding development of credit institutions. Buchinsky and Polak compared data from the West Yorkshire Deed Register, on building activity, with government debt interest rates and found that by the end of our period, and beginning at the end of the eighteenth century, England had become a single market for capital (Buchinsky and Polak 1993).

The joint-stock company, which was to become the standard mechanism for initial capital investment, started off as a mechanism for enlisting working capital. The most successful early joint-stock company, the East India Company, was first chartered in 1601, with stock being terminable after one year and a single voyage. This was repeated twelve times until 1612. From then until 1642 each round of financing was for a cluster of voyages after which the capital was returned to the shareholders. It was only at this point that the Company adopted a permanent capital structure (Baskin and Miranti 1997, 64). Each company however required an Act of Parliament to establish its separate legal personality, and for largely political reasons, associated with monopolising intermediation of government debt, this ceased after an Act banning further incorporations in 1720.

So the foundations were starting to be laid for mechanisms that would enable investment funds to be made available through capital markets, but the industrial revolution succeeded without making a great deal of use of these.

3.2.4 Capital and Creative Destruction

Perhaps ironically, once the mechanisms were in place to create credit, the need became apparent for a mechanism to destroy it. Credit, like any form of money, is a claim on real assets. From time to time investments will be made that will not pay off, and often in these situations the real assets on which claims can be made will not be available in sufficient quantity to fully repay the debt. In this situation the whole system benefits by recognising this and annulling the claims that exist. Otherwise unsustainable tensions would build up and the system would break down. This is the role of bankruptcy: it is the controlled destruction of money and credit, when the economic value from which it derived its value has been destroyed.

After 1706 businesses, defined as traders with debts of at least £100, were protected from the worst sanctions of the age such as debtors prison. They had to give over all their property to their creditors, and when eighty percent agreed they were released from outstanding obligations. If they repaid forty percent of their debts, they were allowed to retain five percent of their estate. Rather than brand bankrupt traders as dishonest parasites on society, the objective of the new approach was to give them a clean slate on which to start again (Daunton 1995, 249).

3.2.5 Financial Revolution

From this survey it can be seen that effective credit mechanisms were in place and increasingly widely used after the beginning of the eighteenth century. Not only was there Dickson's revolution in public finances, there was also a revolution in private financing, though in the initial stages limited to the provision of working capital.

At this stage of economic development, demands from businesses for fixed capital were relatively low, and were frequently combined with the working capital needs of the business, which given the timescales involved in many transactions were considerable. When the demand for capital investment began to increase, there were problems, due to the restriction on establishing joint-stock companies. The next section describes developments that continued to take place through the eighteenth century around forming companies, which could then seek to attract finance. Once these changes were in place it was possible for large corporations to develop, and develop their own commitment and credit mechanisms, but that story must wait until the next chapter.

In a nutshell, for the first time in history a large body of people had the necessary institutional legal and financial support for sustained economic growth. The industrial revolution was no historical accident.

3.3 Consequences

The previous two sections have set out evidence that suggests there were effectively, if not overnight, nonetheless observable, revolutions in legal contracting and financial markets through to the beginning of the eighteenth century. This section sets out to examine the evidence for a corresponding revolution in economic growth in the subsequent historical era, in the four areas of income growth, firm growth, credit instrument variety and the start of a business cycle. After this the relevance of these to one business, namely that of Boulton and Watt, will be examined, before discussing what conclusions can be drawn from the evidence.

3.3.1 Output Growth

It is perhaps ironic that developments in contract law, and to some extent financial markets, are not regarded as significant to wider social and economic developments, but are well documented and, judging by the relatively small literature, uncontroversial. In contrast, the impact of economic growth is seen

as significant, although concrete evidence is sparse and, possibly because of this sparseness, subject to a considerable amount of controversy. Indeed, it has spawned so much scholarship that the constraints of this project make it impossible to present a thorough review of the literature. Nevertheless, it is possible to summarise the evidence and issues in a relatively straightforward way.

First, what is the observable evidence that there was a growth in output through the eighteenth century? In summary, it is considerable. Contemporary sources document many visible signs of a growth in output. Defoe in his *Tour through the Whole Island of Great Britain* in the early eighteenth century found what he called 'manufactures' in most of the areas he visited and believed that their existence added to local prosperity (Sharpe 1997, 146). Hudson describes a number of studies that, although challenging each other on the details, show an overall shift towards the creation of consumer goods markets. Goods quoted include knitted stockings, knitted caps, cheap earthenware nails, tobacco pipes, lace and ribbon, processed foods and drinks. Alongside this was a substantial rise in the proportion of people living in urban accommodation, and the corresponding rise in purchase of household essentials such as clothes, beer, butter, candles and crockery (Hudson 1992, 176). One particular example, well known at the time, was that of Josiah Wedgwood's pottery business. Although like Böttger a producer of fine china, his focus was like Darby on the mass market, for which his focus was cheaply manufactured goods. His principal innovations however were in his development of marketing techniques such as advertising and distribution channels (Hudson 1992, 174). Other entrepreneurs also became prominent in the 1760s. Although perceived primarily as technical innovators, as in the cases of Guttenberg and Darby, there were considerable problems of financing, organisation and marketing to address. What people like Arkwright and Hargreaves did was not just to invent machines to automate aspects of textile production, but to build sustainable, and for a time very profitable, businesses.

In the second quarter of the eighteenth century river navigation improvements began to be undertaken, which required capital beyond an amount that could be provided by single individuals or small groups of local worthies, and so were formed as joint-stock corporations (Harris 2000, 95). The novelty of this development suggests that undertakings of a more substantial nature than previously were being undertaken, but also that the commercial traffic that would result was expected to make them economically viable within a fairly limited timescale. In 1754 residents of Manchester petitioned Parliament to pass a bill authorising the construction of a new waterway for carrying coal to Manchester and Salford. But the opposition this created was sufficient for the project to be abandoned. With this experience in mind, a similar project near Liverpool was promoted as an improvement project, but the brook that it was proposed to improve was in reality too narrow, and its improvement too impractical to be realistic, so that what was proposed and subsequently undertaken was essentially a new construction. In 1759 the first bill to explicitly authorise the construction of a canal was passed (Harris 2000, 96). Probably for the first time in history, business activity was shaping the environment to greater extent than centralised religious or political-military concerns.

All this evidence presents a clearly observable picture of a growth in output over the course of the eighteenth century. But is it possible to go beyond observable phenomena consistent with output growth and quantify overall output growth? Unfortunately, direct quantification is impossible. It was only in the nineteenth century that extensive statistics were collected. The decennial population census from 1801 and the introduction of the income tax in the 1840s are just two among a number of contemporary sources that provide extensive reliable quantification of many aspects of national life by the middle of the century. This is of course another observable indication of greater overall production.

Nevertheless, by projecting data backwards, and using more restricted datasets such as the limited tax data that was collected and samples of data such as burial records, it has been possible to construct a number of quantitative estimates of eighteenth century English economic activity (Harley 1993, 177). It is generally agreed that output increased most dramatically in three key sectors, namely cotton, metalworking and agriculture. Harley estimates output growth in several industrial sectors (1993, 181). Taking 1841 production levels at 100, in 1770 cotton output was 0.8 and metal 7, then in 1815 cotton was 19 and metal 29.

One series for agricultural output estimates a growth of 0.77 percent annually between 1700 and 1750, rising to 0.80 percent for the next fifty years, and 0.85 for the subsequent half-century (Harley 1993, 190). Much of the growth in agricultural output is said to arise more from capital investment than technical innovation, but the commitment-credit model does not see such a distinction as particularly strong. The enclosure movement can be seen as an intrinsic aspect of this, as the move to private ownership would be a necessary precursor to private investment, and the expected benefits from investment the motivation to undertake the steps to acquire the legal right to enclose. The effect of output growth in agriculture can be observed from the growth in population in Britain from 8.5 million in 1770 to almost 21 million in 1851, when only 20 percent of food was imported (Clark 1993, 230). Growth in agricultural productivity was more significant than these figures would suggest at first sight, because the proportion of people working in the sector fell, from around 50 percent in the mid eighteenth century, to around 25 percent in the mid nineteenth (Clark 1993, 233).

It is clear from these figures that there were dramatic effects on some parts of the economy. Harley argues that the industrial revolution in Britain was essentially a matter of luck on achieving technical breakthroughs in the cotton and iron sectors (Harley 1993, 255). This seems to be a little weak, given that he does recognise that 'the basic character of the economy changed from one

governed by the balance of land and population to one dominated by technological change and capital accumulation' (Harley 1993, 197). He also pointed out that prior to 1800, in general, real wages fell and rose in inverse proportion to population and without any secular change, while subsequently real wages in Europe have steadily increased along with population. Others such as Landes see the steady acceleration of industrial growth from the mid eighteenth to mid nineteenth century as a substantial revolution. Hudson points out that the growth that clearly occurred in total factor productivity from 1760 to 1830 probably took place around 40 years after an acceleration in the rate of new patents and patentable invention. To the extent that this is the case, it confirms even more strongly the commitment-credit model as a useful microeconomic theory of growth.

3.3.2 Firm Growth

While studies of the overall level of output of the economy are relatively abundant, recent assessments of changes at the microeconomic level appear to be almost non-existent. This is perhaps a result of the neoclassical dogma in which non-price mediated organisation of production is taken to be dysfunctional. However, while overall assessments of organisational developments are absent, it is possible to observe the effects of the growth of larger firms from a number of angles.

The first observable development consistent with a growth in the size of economic units of organisation, is what in many ways is the defining characteristic of the industrial revolution, namely the creation of the factory system. Sharpe dates the birth of the modern factory to 1724, when John Lombe opened his silk mill at Derby. Driven by waterpower, it employed 300 women and children in two shifts through a 24-hour day, though it failed financially in the next decade (Sharpe 1997, 149). Richard Arkwright developed his spinning machine around 1768, and formed a number of partnerships to provide the capital to develop it. He patented it in 1769, and shortly afterwards set up a mill at Cromford in Derbyshire. By 1782 he

claimed to employ over 5000 people (Baines 1835, 32). Zuboff in a study primarily focused on the introduction of computerisation to businesses in the 1980s, compares this with evidence of the introduction of the factory system in eighteenth century England, and finds many of the same issues needing to be addressed. In many ways the technical problems can be the easiest to overcome. The organisational and managerial challenges are at least as formidable (Zuboff 1989).

Another observable development is in the legal environment. A growth in firms would create a growth in demand for resolving the disputes that inevitably arise when people work together for an extended period of time. Resolving these disputes effectively requires a body of legislation and legal doctrine dealing with the issues. The eighteenth and early nineteenth century saw a wide variety of possible approaches, before an enduring solution was found in the limited liability company.

The Bubble Act of 1720 is often seen as barrier to the use of the joint-stock form, but in fact it only confirmed the existing illegality of formation without Parliamentary sanction. Rather than a reaction to the excesses of the South Sea bubble of that year, as has sometimes been supposed, this aspect of the Act was a manoeuvre by South Sea Company directors to secure its position during the investment bubble of that year, by making investors wary of other stocks and so more likely to put money into their company's stock.

Given these complications, and the relatively small size of early industrial companies, the joint-stock approach was unpopular because of its cost and complexity. However, the construction of transport infrastructure was one driving force for legal developments in this area. Canals were initially funded through local government or by wealthy landowners, such as the Duke of Bridgewater. But the scale of the undertakings led to projects being incorporated as joint-stock companies from 1766 (Harris 2000, 99). Railways in many ways followed the pattern of canals. In both these cases

Parliamentary approval was necessary, even without joint-stock organisation, because it was the only effective way to ensure that no landowner could hold the company to ransom, because an Act could give the company the right to purchase the land it needed (Harris 2000, 282).

Another early user of the joint-stock form was the insurance industry, which is by nature capital concentrating. Indeed, the granting of a joint-stock insurance monopoly was the purpose of the rest of what has become known as the Bubble Act, and was probably the more significant aspect at the time of its inception. Initially, insurance firms took a very diverse range of organisational forms including monopolistic corporations, incorporated mutuals, unincorporated joint-stock companies, unincorporated mutuals, and a syndicate of individual underwriters, the latter being Lloyds. There were few developments from 1720 until the 1760s, after which many new firms were promoted, quite a few fraudulently. Nevertheless some sustainable businesses were created, including several based in the regions outside London (Harris 2000, 102).

Once the possibility of a standardised incorporation procedure became apparent, one argument against it was that it would diminish property rights. Specific parliamentary approval allowed the use of lobbying by vested interests to create a block, or negotiate compensation in some form, to prevent damage to existing property rights. Harris quotes insurance as a market where this view was particularly prevalent (Harris 2000, 135).

Around the beginning of the nineteenth century developments were becoming more rapid. Harris argues that the rise of limited liability was not monolithic, but by the start of the century it was taken to be an element of Parliamentary incorporation (Harris 2000, 129). As well as transport and insurance companies, the other sectors where joint-stock incorporation was applied were water and gas supply, the latter from 1815, and docks and mines. The latter are probably the only ones not requiring Parliamentary sanction, to

prevent hold up by individual property owners, to construct their infrastructure. Presumably it was their initial scale, and the highly visible and irremovable nature of the investment, that made incorporation a beneficial investment for their promoters. The other sector affected was banking, as mentioned briefly in a previous section, and although incorporation was possible after the Bank of England lost its monopoly on joint stock banking in 1826, banks were subject to their own regulation.

Despite objections, the need for routine joint-stock creation was clear. The 1844 Companies Act introduced a standard process for automatically forming a joint-stock company. But Gladstone, who as a young secretary of state piloted the legislation through Parliament, avoided limited liability as too contentious. This was finally introduced in the 1856 Act.

The final area where evidence of the growth of firm size is readily available is in studies of working class life and early industrial relations. Indeed, in many respects the turn of the nineteenth century saw the beginnings of what became class-consciousness. The way this was realised was to some extent fragmented. Nevertheless, developments were widespread enough for a sustainable argument for the beginnings of class-consciousness to be presented by a number of historians (Hudson 1992, 204). Alongside this, more direct evidence from 1802 is the first Factory Act, which primarily limited child labour. While child labour would have been far from a novelty, the factory system had clearly become a national phenomenon.

3.3.3 Financial Structure

The growing size and sophistication of firms in eighteenth century England, and particularly the substantial capital requirements of canal and railway companies, led to a growing variety in the financial mechanisms available. Again, recent studies do not give much attention to these developments. Nevertheless, several mentions are made, even if just in passing, so some brief details of evidence can be sketched out. In particular, a number of Acts of

Parliament deal with these issues, so they were treated as significant at the time.

Controversy over the questionable nature of share trading was ongoing throughout the eighteenth century, particularly after the collapse of the South Sea bubble in 1720. Nevertheless, trading of the joint-stock vehicles for government debt, including the South Sea Company as well as the East India and Bank of England companies carried on. The twice-weekly *Course of the Exchange* was founded in 1697 and published continuously through the eighteenth century, under a number of publishers, until in 1803 it came out under the authority of the Stock Exchange itself. However until 1810 it only regularly published prices on the three moneyed joint-stock companies (Harris 2000, 120). So, while trading in the other joint stock companies, primarily the insurance companies until the arrival of canal stocks, would no doubt have taken place, it was not a primary concern of the London business community. It is an interesting aside that an Act of 1735 prohibited trading in stock not actually held, thus nipping in the bud the early legitimating of derivative trading (Kindleberger 1993, 192).

While London finance played a role in financing canals and railways, most funding was found closer to the developments themselves, most of which were in the north of England. Up until 1845 only one London banker, namely Glyn, played a prominent role (Kindleberger 1993, 194).

Alongside straightforward equity investment, debt instruments were also used. Debentures were invented during the eighteenth century and were loans made to companies in return for a guaranteed rate of interest. They were believed to be more secure than stock, and could be backed by security in land or other assets held by the company. One security device was to only issue debentures up to the unpaid portion of share capital, so that in theory shareholders would be liable to debenture holders in the case of bankruptcy. In other industries the practice later developed of leaving a proportion of

each share unpaid as security for bank loans or as a reserve that could be called on in the case of accident liability (Kindleberger 1993, 197). An 1836 Parliamentary standing order limited railway loans to one third of authorised capital, with borrowing forbidden until half of the share capital had been paid up (Baskin and Miranti 1997, 146).

Partly because debentures were not, in practice, substantially less risky than equity, the preferred stock was developed for canal companies after 1800 (Harris 2000, 227). It was initially used to provide capital to complete the construction of financially troubled projects. This financial security also did not automatically dilute the ownership of the original stockholders. It provided a guaranteed dividend rate, but if dividends fell into arrears the company would not be forced into receivership. By 1849, 66 percent of total share capital in British railways took this form (Baskin and Miranti 1997, 152).

3.3.4 Business Cycle

In many ways it is curious to suggest that the business cycle is a consequence of the commitment-credit model. Sustainable growth was the novelty of the eighteenth century. Previous centuries had not been static, but cyclical. Obvious cycles described previously include, on the largest scale, the end of the Roman Empire and the Black Death, and the subsequent recovery of the European economy from both. This begs the question, how can the cycles due to shifts in intertemporal expectation be distinguished from other cycles, and indeed are they in any meaningful way different from other cycles? The answer is that the evidence suggests that financial expectation driven cycles did become more significant than other sources of cycles, what are known in the neoclassical literature as exogenous shocks.

One phenomenon already noted in the development of the credit market was the development of a national market in the course of the eighteenth century. The synchronisation of interest rates was used as the test for this, so to some

extent it assumed the existence of cycles, but it is significant that synchronisation did shift in the period under consideration.

Mirowski carried out an extensive examination of the evidence of business cycles, albeit in a very different theoretical context. A number of his observations support the hypothesis that the economic cycle was becoming endogenised in the course of the century. His examination starts at the end of the seventeenth century, and there is evidence that cycles were correlated with wars at the beginning of the period under consideration. However, the boom of the early 1730s, measured by the rate of profit, was not directly linked to a war (Mirowski 1985, 235). Furthermore, a chi-squared causality test for likely cause and effect relationships between a number of variables, including temperature, for its effect on crop yields, exports to Asia, GNP and share prices finds only two relationships to be significant at a five percent level. For the period 1755-1826 exports to Asia predict share prices, and for the period 1830-1875 GNP predicts share prices. For the early period East India company shares were a large proportion of the share index, and as it still conducted a large amount of trade, in conjunction with its role in the management of the national debt, this provides an explanation of the first observation. While GNP was not used directly for the earlier period, a profit index was used as a proxy. This makes it possible to observe the shift to financial market performance following output growth, which is an endogenous economic relationship of the type predicted by the commitment-credit model (Mirowski 1985, 263-264).

Another cyclical phenomenon, market manias and panics, are a well-attested historical observation. Well-documented examples with widespread impacts include the Dutch tulip mania of the seventeenth century, and the South Sea bubble in England, and John Law's financial schemes in France, both in 1720. Similar cycles occurred in England towards the end of the eighteenth century in connection with canal building. Parliament passed 122 Acts for canals in the years between 1755-1814, most in two waves. The first was in the late

1760s and early 1770s and the second in the early 1790s. The difference in this case was that much of the money, though by no means all, went into genuine investments, rather than simply fuelling further price rises until the supply of new money dried up (Harris 2000, 99).

The nineteenth century saw more booms associated with the creation of joint-stock companies, including 1807, and 1825 (Harris 2000, 202, 218). Trumping canal speculation, as it did other aspects of canal building, was the railway-building boom of the mid nineteenth century, and this too created speculative manias and panics. Kostal cites the railway panic of 1845 as an event from which Victorian railways never really recovered (Kostal 1994, 360), although it did lead to the building of 5000 miles of track (Morgan 1984, 617).

One additional source of evidence is a recent chronology of British history (Morgan 1984, 616-617). This lists no commercial booms before 1799, then a regular series in 1799-1801, 1809-1810, 1815-17, 1824, 1835-6. Economic cycles were becoming important events in the life of the nation, along with output growth, firm growth and a growth in the variety of financing. One way to see how these all worked together is to consider the progress of one business.

3.3.5 Boulton and Watt

James Watt was born in 1736, and trained as an instrument maker in Glasgow and London. Following early employment at the University of Glasgow, he went into partnership with John Craig, an architect who provided the initial capital, making and repairing a range of musical and scientific instruments. One of the projects Watt undertook for the university was the repair of a model of Newcomen's steam engine. He was intrigued by its inefficiency, and discussed the problem with Joseph Black, a professor at the university. Black explained the problem in terms of his own theory of latent heat. Using this information Watt was able to devise a mechanism to make the engine far

more efficient. Fortunately, machining techniques were becoming available that would make his ideas practicable.

However, in 1765 Craig died, and Watt had to pay back most of the capital in the business to his trustees. Black was able to put Watt in contact with John Roebuck, an industrial chemist who had established a successful steelworks. He was also attempting to exploit deep coal seams, and was having problems because his Newcomen engine was not properly removing the water entering the workings. Roebuck was interested but sceptical, so would provide only limited support. With the first canal boom underway, Watt took up surveying and civil engineering to make ends meet. By 1768 he had constructed a model that worked effectively enough for Roebuck to finance further work, pay off his debts, and pay the costs of patenting the invention, in return for a two thirds share in the patent (Davenport 1989, 8-13).

On the way back from a visit to London in connection with the patenting process, Watt stayed with a friend in Birmingham and was introduced to Matthew Boulton. Born in 1728, Boulton was already a very successful businessman. His father was a 'toymaker,' whose principal product was buckles, which were used on shoes and knee breeches. It has even been suggested that they were exported to France and then re-imported as the 'latest thing' from the fashion leading French (Dickinson 1936, 25-28).

Boulton had taken over his father's business by 1760, and was also representing the whole of the trade for Birmingham, Wolverhampton and Warwick in a petition to Parliament. By his testimony, around 8000 people were employed across the region in the trade at that time (Dickinson 1936, 31). Over the next few years he established a successful manufacturing operation in Soho, just outside Birmingham, producing steel jewellery and buttons, and employed 600 people by 1765 (Dickinson 1936, 48).

By 1766 Boulton was looking for a solution to a problem of lack of power from his waterwheel, particularly in the summer. One solution he was

investigating was to use a steam engine to pump water back up after it had turned the wheel (Dickinson 1936, 75). Meanwhile, Boulton had already made the acquaintance of Roebuck, who had even made him an offer of partnership in his coalmining venture, which was declined. Boulton was however very interested by the steam engine, which seems to have spurred on Roebuck, and although he offered Boulton an involvement, it was very limited, and Boulton again declined.

1772 saw problematic business conditions, and a tightening of credit, and Roebuck became insolvent. Boulton, with Watt's support, took over the engine patent in 1773 in consideration of a debt of £1200 owed to Boulton from his previous business dealings with Roebuck. Once Watt had completed his last canal survey he moved to Birmingham in 1774 (Dickinson 1936, 84).

Given that patents only had a life of 14 years, by 1774 there were only 8 years remaining on Watt's patent, and it was clear that this would not be enough time in which to make a return on the considerable investment still required. It was therefore resolved, and with some effort achieved, to petition Parliament for an extension of 25 years to the patent. The two men entered partnership, with Boulton supplying the necessary funding in return for two thirds of the profits, and Watt providing his share of the patent and his services in return for an annual salary of £300 and one third of the profits (Dickinson 1936, 86).

Watt continued to work on a prototype, which continued to show that his ideas were practicable. He wanted to develop a larger experimental version, but Boulton persuaded him to move straight to commercial construction. Two contracts were obtained, one for a colliery in nearby Tipton and another for John Wilkinson the ironmaster, and competitor to the still thriving Coalbrookdale company. Wilkinson was also employed to manufacture the cylinder bore using a new process he had developed, originally for drilling cannon bores. The business arrangements entered into for these and

subsequent engines were more those of consulting engineers than manufacturers. The client would pay for all the costs of the parts and construction. Boulton and Watt would supply detailed drawings, some key parts, and an engineer to supervise the construction; other parts could be supplied as the customer chose. The income for Boulton and Watt would come from one third of the coal saved in the use of the engine.

Cornwall with its tin and copper mines became the biggest early market. By 1780 the firm had built forty engines, of which twenty were in Cornwall. Because direct comparisons were not directly possible, the contracts actually specified payment based on coal consumed and water pumped. A tax rebate on coal used in pumping water from mines meant that this was reliably measured, and Watt developed a counter to measure the number of strokes made, which allowed the water pumped to be measured. Nevertheless, the payment arrangements, at least in part due to their complexity, were proving problematic, and Boulton negotiated a fixed annual premium depending on each engine's size. Boulton's other business lines were also expanding at this time, and he was able to use this income as security against capital invested in his growing business, a relatively unusual practice at this time. As has been observed, the normal practice would be to enter into partnership with a capital provider, but Boulton already having several partners was reluctant to enter into more, each of which would incur a joint and several liability (Dickinson 1936, 97). By 1780 it was clear that the toymaking business was making a loss, and income from the steam business was not yet substantial. It was 1785 before the business was financially secure (Dickinson 1936, 112).

In the meantime, Boulton was conscious of the limited size of the Cornish market, and consequently of the need for diversification. He encouraged Watt to turn his attention to applying steam power to rotating motion, which would provide an alternative to waterpower, and for which he could foresee considerable demand. The obvious route was to use a crankshaft, but this had, albeit somewhat dubiously, been patented immediately prior to this, and

to avoid confrontation Watt devised a number of alternatives, the one that came to be used being the 'sun and planet' gear. The charging mechanism was slightly different for rotating engines, and was based on the equivalent horsepower of the machine, the annual charge being £6.6s in London and £5 in the provinces, per horsepower. Wilkinson took the first in 1783 to power a tilt hammer.

Another early application was the Albion flourmill. The promoters, including Bouton and Watt, applied for joint-stock incorporation, but mainly due to opposition from existing mill owners, this was rejected (Harris 2000, 178). The Albion mill burned down after five years production, and was not rebuilt. Boulton and Watt lost several thousand pounds from their initial investment, but one innovation was developed as a result. Boulton observed that a rotating sphere mechanism was used to lower the rotating upper wheel of the mill as its speed increased, and vice versa. He communicated this to Watt who applied the mechanism in what was possibly the first example of an automatic feedback mechanism. For his rotating engine, as the speed of rotation started to increase, a valve operated to diminish the supply of steam to the cylinder, thus reducing the pressure and the speed of rotation. Although this was a remarkable advance over its application in the flourmill, its superficial similarity probably discouraged Watt from applying to patent his application (Dickinson 1936, 126).

The shift to rotary engines meant contracting for engines with firms that did not have existing sources for supplying machinery, and the Soho factory was increasingly called upon to provide many of the parts. By 1793 approximately fifty percent of engine parts were manufactured there. It was also apparent, particularly to Boulton, that with the expiry of the patent in 1800 the firm's competitive advantage as providers of drawings and expertise would not be much of a differentiator. To this end, in 1796 a new plant was built not far away, but adjacent to the canal network, and set up legally as a partnership between Boulton and Watt and their sons (Dickinson 1936, 168). Watt senior

gave his share to his surviving son, and when Boulton died in 1809 the two sons held equal shares in the business, which continued to be regarded as the market leader in steam engine production for many years (Dickinson 1936, 171).

3.3.6 Conclusions

The consequences of the industrial revolution in England were deep and profound, and so far irreversible. By the end of the 1850s most of the economic environment we now recognise was in place. In 1851 the Great Exhibition brought mass crowds to London, many on the railway. Also in 1851, the Amalgamated Society of Engineers formed, the first 'New Model Trade Union.' The 1852 Patent Act brought intellectual property law into the modern age, and the 1856 Limited Liability Act brought into place most of the modern features of corporate governance. In practically every indicator of human life the changes that came about in this period were also enormously beneficial.

The previous revolutions in legal contracting and financial markets predicted by the commitment-credit theory, while not so well known, were just as observable. So, not only do historical observations confirm the commitment-credit theory, the theory also provides a clear and concise explanation of important historical developments.

Chapter Four

MANAGERIAL REVOLUTION IN THE USA 1850-1950

Britain was the wealthiest and most productive country in the world right through the nineteenth century; but it was overtaken at the end of the century by its former colony, the United States of America. This chapter focuses on that shift, to test another prediction of the commitment-credit model, namely that the rise of the multidivisional company, which was the most visible economic difference between the US and the rest of the world in the early twentieth century, was preceded by fundamental shifts in commitment and credit mechanisms; and that its consequences included growth as envisaged by the model. As in the previous chapter, the dates are only indicative. Nevertheless, before 1850 there was little distinction between economic life in the US and Britain, and in many ways the US was behind Britain. While by 1950 the US was clearly the dominant economic superpower, probably to a greater extent than it could ever have been able to maintain, although subsequent developments will not be discussed here.

Both of the previous chapters, dealing with pre-modern Europe and the industrial revolution in England, have had to cope with limited primary data; but from the second half of the nineteenth century extensive data becomes available. The problem with data for most of this period is that although primary sources exist in abundance, the secondary sources are fewer than they are for the industrial revolution. To some extent, it can be argued that there are only fewer recent works compared to the amount of data available. But it is also likely that this period is often perceived to hold less interest.

Much of the study of modern business tends to rely less on general theories than on the analysis of a variety of cases. The neoclassical model that dominates contemporary economics, in contrast, ignores the key aspects of

the transformation, such as market concentration, as well as the fundamental observation this thesis sets out to explain, the increased level of innovation, and consequent increase in growth rates. So when historians or economists interested in testing the neoclassical model look for historical epochs where the model may help understand changes, this is not an era that has much to offer. So by explaining many observations from this epoch, the commitment-credit model again offers an improvement over neoclassical theory.

One work in particular dominates the secondary historical literature, Alfred Chandler's *The Visible Hand*, and is the most widely used source relied on here. Nevertheless, other work does exist and is used here to ensure that a variety of narratives are used in testing the hypotheses. As before, because hypotheses are being tested, rather than a narrative being set out, and the role of other narratives here is in making the primary evidence accessible, the use of a relatively small number of sources, and little claim to comprehensiveness, is not the weakness it would be if the objective were simply the construction of an explanatory narrative.

There is overwhelming evidence that a major transformation did take place in the nature of economic activity at the end of the nineteenth and the beginning of the twentieth centuries. Centrally directed business organisations of a size never seen before came into existence, and organisations of this type grew in influence as the century progressed. The commitment-credit model suggests that this should have been preceded by significant changes in the commitment and credit mechanisms available to the managers who exercised overall control of these enterprises. The model furthermore suggests that this growth in size will have had four aspects, first that the economic output of these organisations will have increased, secondly that observable distinctions at an intermediate level between the overall organisation and the individual people who made it up will have grown in size, thirdly that the variety of mechanisms by which credit was provided to allow investment will have increased, and finally that within the overall growth in output there will have

been an observable cycle in the of rate of change. This chapter sets out the evidence that these changes can be observed.

4.1 Commitment

It is the principles of financial accounting that provide the mechanism through which commitments are made within large organisations, and between large organisations and their investors and creditors. Very large private organisations first developed in an economy governed by the rule of law, and in particular with contract law, broadly understood, as set out in the previous chapter. But within these large organisations, it was not legalistic contracting that formed the basis of most agreements, but budgets and financial reports. Much as with contract law, effective principles only become 'obvious' over a considerable period of time, and other approaches were tried along the way that, although promising at the time, proved to be inappropriate. This section sets out these developments.

Although it was only at the beginning of the twentieth century that the whole conventional framework of modern accounting became established, it is in the commercial maelstrom of the late middle ages that the concepts at the core of the modern conception of accounting have first left a visible trace, particularly in late medieval Italy. The origin of the term accounting is in the concept of giving an account to someone who has entrusted one with an office or activity; that is, submitting a report on the extent one had lived up to the commitment one gave in taking on such a responsibility. The origins of commercial accounting must be far earlier, hidden in the mists of time. Some of the earliest records made by human beings concern what would now be labelled as accounting information. When the concept of number and of writing emerged at a similar time in the evolution of humanity they were probably intrinsically connected with recording economic activity. The word itself is very ancient; and as far back as the thirteenth century there was an 'action of account' in English law (Harris 2000, 161).

4.1.1 From Bookkeeping to Accounting

Historical certainty begins with the publication in print of Luca Pacioli's *Summa* in 1494 (Previts and Merino 1998, 4). Some sources indicate that the double entry concept from which accounting builds was employed as early as the fourteenth century, in Genoa and Florence (Previts and Merino 1998, 2, 3). The double entry convention recognises the essential truth that every commercial transaction has two sides, a giving and a receiving; as well as providing an effective check, even more important in pre-computer ages, that the entry has been recorded correctly.

While double entry ledgers in themselves solved the problem of maintaining records of complex transactions between a firm and its customers and suppliers, they did not in themselves provide a way of understanding what was going on inside the firm. Clearly, for relatively small owner-managed entities this was not a significant problem. An intimate knowledge of their business, gained through daily experience dealing with people inside and outside the business, making purchases, and seeing activity and output in the workshop, would provide them with sufficient data to enable them to manage the business effectively.

Nevertheless, there is evidence that for as long as firms have used double entry ledgers to keep track of their external commercial relationships, they have also kept records to help ascertain the economic relationships between different elements within the business. Garner provides an extensive number of examples of developments from the middle ages until the 1920s. Early examples include Genoese ships' accounts. From the thirteenth century many maritime towns in Italy required a seagoing ship to employ a scribe. Their careful records became public documents so, as it became common for several people to sponsor a voyage, there was a clear basis on which disputes could be resolved. There is even some evidence of the use of relatively sophisticated concepts such as depreciation, by which the cost of something used over a period of time is spread over that time (Garner 1954, 3, 4).

In the fourteenth century the Del Bene wool manufacturing firm used two sets of books that, although not double entry, separated trading from manufacturing activities (Garner 1954, 15). Towards the end of the century, records of Niccolo di Pero, a dyer in Prato, compute a cost of production, and then adjust it to allow for the revenue from two by-products which had a market value, namely filling and warp yarn (Garner 1954, 19).

In the sixteenth and seventeenth centuries the Fugger family developed a structured set of accounts for mines they owned, and for linked activities such as a foundry. Over the same time period the Medici family produced silk and wool, and many ledgers, journals and similar records are extant. An extensive system of double entry records enabled calculations to be made to calculate a cost of finished cloth manufactured over a particular period of time, based on the costs of wool, dyestuffs and manufacturing costs such as workmen's wages, oil, combs cards and teasels (Garner 1954, 9). Contemporary accounting texts do not match the sophistication the Medici achieved in practice, although some manuscripts dealing with textile manufacture do show a greater sophistication than the accounting texts. Some depreciation is taken into account, but overall capital costs such as machinery are not factored in. In large part this is understandable given that, as the previous chapter shows, even in the industrial revolution in England, investment capital costs were not a significant element of manufacturing costs (Garner 1954, 15). While the examples described here do show a considerable level of sophistication, it is not clear to what extent the practices were widespread (Garner 1954, 25).

From the very beginning of its European settlement, commerce was a fundamental part of life in what were to become the United States of America. As such, commercial arithmetic, sometimes called casting accounts, was a key element of formal education (Previts and Merino 1998, 21). The heart of any mercantile establishment was the counting house, and this was where transactions were recorded, providing both an authoritative record and

a mechanism for determining from time to time the capital position of the business (Previts and Merino 1998, 48). By the end of the eighteenth century textbooks on the 'modern Italian method' were being written and published in the US, and establishments for teaching it began to develop (Previts and Merino 1998, 52). Some early teachers were as much entrepreneurs in their own right. Between 1853 and 1866 Bryant and Stratton developed a chain of more than fifty commercial colleges (Previts and Merino 1998, 76). In practice, most learning continued to be as it had been for centuries, as an apprentice clerk in an existing establishment.

It was only with the work of Thomas Jones, including his textbook *Bookkeeping & Accountantship* published in 1859, that the foundation of teaching moved away from ledger balances, and rules for specific types of account entry, towards a focus on the financial statements as the objective towards which financial transactions were recorded (Previts and Merino 1998, 78). This signals the shift from bookkeeping to financial accounting.

4.1.2 Accounting for Management Control

The industrial revolution had magnified many times over the problems early cost accounts had wrestled with. More complex processes and increasing sunk costs made the 'cost' of a product more difficult to define precisely. Nevertheless, developments continued. One major difficulty in assessing the evidence is that cost accounting, even at the beginning of the twentieth century, was seen as a source of competitive advantage (Garner 1954, 30). So whereas well-kept systematic trading accounts would probably help build confidence with customers and suppliers, as well as being widely recognised as a necessity for all but the smallest merchant, for a manufacturer the benefit of making accounting information available was far less clear.

There is evidence that the introduction of formal systematic accounting for management control was primarily a response to increases in working capital (Levenstein 1998, 29). And one commentator on nineteenth century

glassmaking in the United States attributed frequent bankruptcies to a lack of reliable cost accounting systems (Garner 1954, 76). A range of studies suggest that through to the 1870s the introduction of cost accounting systems was a strategic response to the need to control labour and material costs in the face of competitive pressure. New England textile mills wanted to be able to determine the extent to which materials and labour were wasted in the production process. The paper industry focused more on monitoring and controlling raw material and finished product inventories. In one well known case, Lyman Mills in Holyoke, Western Massachusetts, while transactions were recorded and data exchanged between mill and head office every day, the actual calculations were only carried out every six months (Johnson and Kaplan 1987, 25). The effect of having this information available was probably a device for motivating managers, it was unlikely to be available promptly enough to provide an effective input into planning decisions (Levenstein 1998, 31).

Formal internal cost control systems also show evidence of having been introduced primarily in situations where there was a separation between management and ownership. In the case of New England textile mills this separation was geographical. The owners had offices in Boston, where they negotiated commercial arrangements for purchases of materials and sales of finished goods, while the mills themselves were inland, managed by agents.

Accounting innovation in the cut nail industry followed the sale of several companies from groups of partners to groups of stockholders outside the industry. A novel approach was taken in the 1850s by firms in the Ohio River Valley, where a standard cost accounting system was introduced to allow firms to compare production performance (Levenstein 1998, 31). The latter was clearly an enlightened attempt to shift competition away from destructive price competition, to competition on quality and service.

The inventory monitoring of the paper industry, which retained direct management by the owners, was carried out simply through elaborating the standard mercantile accounting system with the use of detailed and subdivided inventory accounts (Levenstein 1998, 32).

There is evidence from the 1860s that some cost data was collected for the direct purpose of decision-making. One example is of data analysed to help in comparing the impact of different power sources in the paper industry, and later on choosing between cloth and wood as the basic raw material. There is evidence of a transition period in the textile industry. Managers at Lyman Mills used costing information to help decide issues of minor machine modifications, while in other mills only the agent, the highest-level local manager, had access to the figures, while the 'overseer of repairs' on the factory floor made technical decisions.

Increasing product differentiation resulted in different competition dynamics in different markets. Where products were uniform there was often an industry association that set a standard cost calculation, to help prevent destructive price competition. For special-order driven industries, often employing a tender based selling process, the incentive was for individual firms to understand their underlying cost structure as accurately as possible. Where this was not possible, establishing a relatively accurate variable cost for a product would set a minimum acceptable price. Sophistication developed even to the extent that by the end of the century many firms were responding to slowdowns by reducing output rather than prices (Levenstein 1998, 34-5).

The next problem, to which some of the answer was found in cost accounting, was long-term capital investment decision-making. Early in the twentieth century DuPont Chemical is the first company for which evidence exists that a rate of return of current processes was calculated, based on the profit earned for the capital invested in a particular factory or product (Levenstein 1998, 36). However, this calculation needs a figure for the capital

invested, and this is not as straightforward as it may seem at first sight. But before delving into the details of measuring capital investment, other developments largely in parallel with cost accounting need to be understood.

4.1.3 Railway Accounting

The railroads were for much of the nineteenth century the largest and most complex business enterprises that the world had ever seen. A combination of commitments to repay large initial investments and tough competition meant that poor decision making would have serious consequences, so considerable investments were made in collecting and processing the complex mass of data related to road operations. Chandler identifies this as the point at which bookkeeping became accounting (1977, 109). However, capital and operational accounting remained largely separate, although even after a century of subsequent developments a distinction endures, suggesting that there are considerable difficulties in reconciling the two.

As by far the largest integrated railroad for many years, the Pennsylvania played a leading role in many developments. As early as 1851, the company's annual report showed information such as the number of passengers each month embarking from each station, as well as local and through tonnage passing each way-station. While even the largest textile mill had four or five sets of accounts to integrate, by 1857 the Pennsylvania had one hundred and forty four. While mill accounts were compiled semi-annually, for the railroad it was monthly, presented to the directors by the fifteenth of the following month. Comparison between railroads was not simply between profits or losses, but focused on the 'operating ratio,' comparing earnings with the volume of business (Chandler 1977, 110).

So much for collecting and aggregating the raw data: the problem remained of how to deal with the fact that because much of the cost of the road had been incurred before the first train ran, it was difficult to judge the overall economic viability of the business. A crucial distinction was made between

the construction, or capital, account and the operating account. The operating account was made up of expenditure that was essentially consumed immediately, while the capital account showed expenditure of enduring value. Nominally, it was this enduring value that gave a value to the business. Although fairly clear conceptually, in practice fine distinctions often had to be made. Sometimes there were strong incentives to misallocate expenditure. For example, charging operating expenses to the capital account would have the affect of both increasing operating profits and adding to the book value of the business. Alternatively, the opposite tendency would make operating cost look higher, probably leading to higher prices, and while the book value would be lower, the expected future income stream would be higher (Chandler 1977, 111).

Once a road was in operation, there was in some ways an even more difficult accounting problem; namely how to take account of the cost of initial construction in relation to the income received. Related to this was the question of how to deal with renewals and repairs. On the one hand, these were a cost necessary for the continued operating of the road. On the other, they could be considered as adding to the capital value of the business, particularly where they involved upgrades and improvements. Initially the widely used practice of putting funds into contingency or surplus accounts was employed. From time to time, generally when profits were strong, the initial cost of plant and equipment would be written down. However, in 1855 the Pennsylvania explained that it had adopted what became known as 'renewal accounting'. Repairs and even direct replacement of existing plant was recorded as an expense, and only clearly additional assets or improvements, but only the additional value of the improvement over the original cost, were added to the construction account (Chandler 1977, 111-112).

Along with these developments in measurement, the potential for inaccuracies for any reason, accidental or voluntary, was also apparent. So it

became common practice to employ consulting accountants as auditors. Previts and Merino quote in full one letter addressed to the President and Directors of the Western Rail Road from William Ritchie in 1855. Five brief paragraphs set out what he did, checking the consistency within and between the different sets of books in use, and his findings, namely that everything is in order. An 1857 annual report for the Boston and Worcester Railroad included a four-page report, including recommendations for improving the system (Previts and Merino 1998, 85-86).

The 1860s, at least after the Civil War, saw attention switch from capital to cost accounting. The cost per ton-mile was already well established as the basic operational measure in railway companies. A man named Albert Fink was largely responsible for this next phase of developments. Initially trained as a civil engineer and bridge builder, he worked on the Baltimore & Ohio line before joining the Louisville & Nashville, rising to become senior vice president in 1869. An important precursor to the system he introduced was to re-categorise existing accounts according to the nature of their costs, rather than according to the organisational units in which the function was carried out. Fifty-eight distinct expense accounts were established, in three categories. Twenty-three accounts of general maintenance and management costs were allocated to maintenance expenses per ton-mile. Eight types of station expenses were distinguished, and twenty-seven types of movement expenses. These three categories, combined with data on interest and construction costs, miles of track, miles and freight weight run, as well as revenue, enabled Fink to establish four components of the total cost per ton-mile, namely movement expenses per ton-mile, station expenses per ton-mile, maintenance of road per ton-mile, and interest per ton-mile (Chandler 1977, 116-119).

Attempts at self-regulation, to prevent destructive price competition, had broken down by the 1880s, and in 1887 President Cleveland signed the Interstate Commerce Act, which established the Interstate Commerce Commission (ICC). Initially this was part of the executive as a branch of the

Interior Department. Railroad rates had to be published and 'reasonable and just.' This also satisfied the big direct customers, the shippers, whose primary concern was parity of rates with their competitors. The commission had three powers, namely compelling witnesses, petitioning a court of law for a determination, and requiring from railroads annual reports using a uniform accounting system. Although the standards specified by the ICC drew on those widely applied within the industry, for the first time, firms found to be failing to use accounting standards were subject to serious penalties.

The administration of President Harrison, which commenced in 1888, created a new type of body when it cut the ties of the ICC to the Interior Department. Although the president continued to appoint members, for seven year terms with no one political affiliation being allowed to dominate, it became the first independent market regulator (Vachal 1993).

4.1.4 Financial Accounting

The benefits of independent confirmation of a corporation's regular financial reports to its investors were always apparent. In the 1870s it was common practice for teams of shareholders to visit corporation headquarters in order to verify reports. While few corporations, outside the railroad industry, continued to exceed ten million dollars in assets, this was to some degree a reasonable approach. However, the 1880s saw a significant increase in the numbers of what were by then becoming known as public accountants. The city directories of New York, Philadelphia and Chicago for example show an increase in the number listed from 81 in 1884 to 322 in 1889 (Previts and Merino 1998, 132, 133). Success was not guaranteed for new entrants. For the same three cities between 1850 and 1899 1,370 different individuals were listed, of whom 662 appeared only once. Nevertheless, a recognised profession was beginning to emerge, and 1882 saw the creation of its first professional organisation, the Institute of Accounts of New York. It required a test of qualification prior to admission, and published a periodical named *Accountics* (Previts and Merino 1998, 135-6).

By the 1890s the Institute of Accounts and a competitive organisation, the American Association of Public Accounts (AAPA), began to seek legal recognition for their profession. Both submitted Bills to the New York state legislature at the beginning of 1895, the AAPA to the State Senate, the Institute to the State Assembly. In March a meeting was called to attempt to reconcile the differences, and a committee was established. The Institute's approach of a recognised title, Certified Public Accountant, was chosen over the Association's proposal for licensing. After a further year of work, from April 17, 1896, the law recognised the profession. Certification was the responsibility of the Regents of the University of the State of New York and, by the end of 1897, 108 professionals were certified on the basis of being in established reputable practice, and 4 were certified through passing the examination (Previts and Merino 1998, 138-141). Other states followed the New York example, Pennsylvania in 1899, Maryland in 1900, although it was the mid 1920s before all states had legislation in place (Previts and Merino 1998, 144).

One early market for the services of independent accountants was the public sector, where they offered scrutiny of public officials to guard against corruption and inefficiency. The introduction of a uniform accounting system was popularly believed to be one way of combating inefficiency at least. The irony was that as these systems were introduced to many municipalities, in the first decade of the twentieth century, it became apparent to the officials involved that uniform systems did not require professional accountants exercising their independent judgement, but technicians trained in the rules, so that particular market for professional services quickly dried up (Previts and Merino 1998, 178-9).

While official employment, at least initially, is perhaps not a surprising consequence of official recognition, in practice the primary effect of this recognition was on conceptual standardisation. For example, what to examine would have been one significant question. This would have been influenced

by practical issues of current concern, and in turn would shape the way those issues were dealt with. Given that the issue of reporting capital investment had already been identified in certain contexts as the central problem, it should come as no surprise that this was in many ways central to the debate around establishing a firm conceptual underpinning of financial reporting. An additional factor was the difficulty, in some ways the most easily recognised, of general price changes. From 1865 to 1897, with a respite between 1879 and 1884, deflation persisted, and then from 1898 to 1919 inflation took over (Previts and Merino 1998, 209).

Through to the end of the nineteenth century, in general, the focus of an audit was on the current value of the business in terms of its assets, conventionally set out in terms of a balance sheet. To a large extent this reflected the prevalent view that ownership of shares in a company corresponded to ownership of the assets of that company. This focused attention on the relationship between revenue and capital transactions. Depreciation started to become accepted as the best way of dealing with the relationship between capital and revenue items, where capital is defined as those purchases that were used for more than one accounting period, generally a year, and revenue those that were consumed within that period. Depreciation allocates a proportion of the total cost of an asset used over a number of years to the costs for each of those particular years, until all the original cost is accounted for. This superseded renewal accounting, as developed in the railroad industry, which up until then had been viewed as the best available practice.

The focus on the balance sheet helped to maintain a good degree of confusion over the role of depreciation, even within the profession. To some extent this was because a claim that the balance sheet should provide a valuation of the firm led to debates such as whether current market values or historical costs should be used. In some cases depreciation was treated as a reserve for replacing existing assets, though in general it was understood that

this was a spurious application. However, by the nineteen twenties it was becoming appreciated that the value of a firm was a function of its future earning power, and there was no necessary logical connection with its balance sheet, except as possibly setting a minimum 'break-up' value on the firm, and this would explain why the value of a firm whose shares were actively traded, and so for which an overall valuation could readily be calculated, could be higher than the asset value shown on the balance sheet (Previts and Merino 1998, 218-221).

Depreciation only really become standard practice once it was included in the calculation of corporate excise tax, introduced by the federal government in 1909. Although shortly after it was introduced this tax was declared unconstitutional by the Supreme Court, and only came into effect with the ratification of the sixteenth amendment in 1913 (Previts and Merino 1998, 181).

Given the necessity for applying judgement on issues such as depreciation, the role of the accountant was from the earliest days inherently problematic. One difficulty was the need to precisely express fundamentally unknowable quantities, not least because they depended on an uncertain future; for example the length of time for which a piece of capital equipment would be employed. Another problem was the difficulty of dealing with the sheer quantity of data. Within the firm, or as an advisor to the firm, policies had to be decided upon, and in their role as an independent auditor, the accountant had to be able to make a meaningful determination of the reasonableness of the policies and the extent of their consistent application.

Understandably, there was resistance, and not just from self-interest in remaining well-paid autonomous professionals, to the standardisation that some saw as the solution to these problems. This would just create a different set of problems. For example, when a general rule had to be applied in a particular case, but a different treatment would have been a better way of

reflecting the underlying business situation. And with more rules would come more occasions for mistakes, until they lost their usefulness because the selection of one among a number of rules would become in turn a question of professional judgment.

The solution found to the problems of fundamental uncertainty and a variety of justifiable interpretations, albeit one with its own limitations, was the need for ethical practice among accountants. To a large extent this was seen as being taught during the period of professional experience necessary, in addition to examination success, in becoming a CPA. Debates were ongoing in this area, but in practice the ACPA applied a concept of “acts discreditable to the profession.” However, given the voluntary nature of the professional association this was not a strong deterrent to questionable practice. Because of the opposition of some to any form of self-regulation, names of persons found guilty were not publicised. However, facts of the case and determinations were, so for the vast majority who wished to employ best practice there were guidelines available. As has remained the case however, concerted attempts to distort accounting information are almost impossible to prevent (Previts and Merino 1998, 205-206).

4.1.5 Managerial Accountability

Unlike in Britain, in the US an annual audit was not compulsory for corporations, but was done for the sake of investors, normally by management as a confidence building measure. Clearly, this was most important when a firm needed or anticipated a need for substantial new investment. The great merger wave of the 1890s was a particular catalyst for this. The audit was management’s commitment to the veracity of the information they provided to the investment bankers who provide large injections of capital, the subject of the next section.

The system continued to function through the first quarter of the twentieth century, even with a significant hiatus for the duration of the first world war,

although its impact on the United States was far less than in Europe. Nevertheless, the international nature of the financial markets was brought home by the need felt necessary to close the New York Stock Exchange (NYSE) on July 31st, the day after Russian mobilisation commenced and European stock markets suspended operation. All restrictions were only lifted on April 1st 1915 (Carosso 1970, 193, 195).

In retrospect, the financial markets had got seriously out of hand by the end of the 1920s, and the Wall Street Crash, particularly 'black Thursday' of 24th October 1929, threatened to destroy confidence in the whole system. By December 1931 the US Senate had authorized an investigation into the NYSE, which became known as the Grey-Pecora Investigation, after the two lawyers who acted as council to the committee and carried out much of the investigation. By the time its hearings came to an end in December 1933 a catalogue of deception and abuse of trust by both corporate managers and bankers had been brought to light. Even before it had completed its investigations, significant new legislation had come in to force (Carosso 1970, 348-351).

Two acts in consecutive years, the Securities Act of 1933 and the Securities Exchange Act of 1934, proved to be particularly important. The Glass-Steagall Banking Act of 1933, in mandating a separation between investment and commercial banking also had a significant impact on the industry, although in hindsight this innovation has not stood the test of time. More than just legislation, as existed in Britain to some extent already, the two securities acts established an effective regulatory mechanism for supervising the operation of financial markets, and in particular defined the disclosures that corporations were required to make. In many ways this built on the regulatory frameworks previously established to effectively supervise transport and federal trade. Indeed, initially the Federal Trade Commission (FTC), established in 1915, was given the role of regulator, though it was quickly relieved of the

considerable burden that the new role required with the establishment of the Securities and Exchange Commission, widely known as the SEC.

The Securities Act was a key plank of President Franklin D. Roosevelt's 'New Deal,' and was one of the actions taken in the first hundred days of his administration. The great depression was at its worst, and the investment banking business had almost collapsed. New corporate security offerings had gone from almost \$9.4 billion in 1929 to less than \$400 million in 1933 (Carosso 1970, 352). Proposals for reform had been campaigned for since the 1880s, and earlier in the century several states had enacted so called 'blue-sky' laws, named because they were designed to prevent the selling by fraudulent promoters of building plots in the blue sky (Carosso 1970, 156). However, the Securities Act was the first ever at a Federal level, and in contrast to the state laws aimed not at endorsing or limiting particular types of security, but in ensuring full and fair disclosure. The issuer of a new security was required to provide detailed information, along with a fee of one hundredth of one percent of the total offering price to the FTC, and most of the information had to be set out in a prospectus to prospective investors, the form and content of which were carefully defined. Registration statements did not become effective until 20 days after filing, providing both a 'cooling off period' and time for analysts and journalists to digest and comment on the data. Failure to abide by the conditions imposed wide ranging civil liabilities, as well as criminal liabilities of a \$5,000 fine and five years imprisonment (Carosso 1970, 356-9). In addition, the FTC was given the authority to prescribe accounting standards in very explicit terms, such as policies for differentiating investment and operating income and for determining a standard depreciation policy (Previts and Merino 1998, 273).

The Securities Exchange Act of 1934 extended and in some relatively minor ways amended the Securities Act of 1933. While the Securities Act primarily concerned new offerings, the new Act extended federal supervision to trading on exchanges. It required the registration of exchanges and regulated their

operations and members, and established the SEC to take on this role, as well as that previously given to the FTC in the 1933 Act. The Commission itself was to be made up of five members, no more than three of whom could belong to the same political party, and it was given wide discretionary authority. Corporations whose securities were traded on the exchanges had to register themselves, file regular detailed reports and report the transactions of their officers, directors and major shareholders in the corporation's securities. Among the several small amendments to the 1933 Act, the most significant was the limitation of an underwriter's civil liability to the price of the proportion of the securities they had underwritten (Carosso 1970, 364, 379).

The shift to valuing firms on the basis of their future earning power, rather than the capital previously invested in them, finally became firmly established by the end of the 1930s, when bodies such as the NYSE and the SEC, and CPAs in general, came to the view the income statement as the focal point of accounting (Previts and Merino 1998, 278).

4.1.6 Accounting Revolution

By the end of the nineteenth century the accounting profession was firmly established in the United States. It was another four decades before the conceptual issues on representing the economic condition of a large corporation were resolved in a manner that has largely stood the test of time, but standards were starting to be enforced, sometimes directly such as by the ICC for the railroad industry, and sometimes indirectly through the standards of peer review to which the CPAs hired to perform the regular independent audits required by investors held themselves.

Accounting information is by its very nature historical, and past performance is no guarantee of future returns, but by the end of the nineteenth century managers and investors both knew that managers would be held to account for their performance in a relatively well understood way. To secure the investment necessary to realise the profit opportunities innovation made

possible, managers had to commit their whole enterprise to achieving a level of performance necessary to service the equity and debt instruments they exchanged for the credit, and be able to demonstrate that this was the case.

4.2 Credit

Alongside the rise of management accountability, the end of the nineteenth century also saw sums of money coming together on an unprecedented scale, and available for privately owned businesses to spend on long-term investment. Projects that previously could not have been contemplated, because they required such a large initial investment, began to be considered as a matter of routine. This section sets out these developments in detail. Again, as with the previous chapter, this is not meant to suggest that developments in credit were separate from those in commitment devices; in this era they were if anything more closely interrelated than previously. However, as well as easing the process of exposition, this division also emphasises the separate logical nature of the developments, from the perspective of the commitment-credit theory.

The early development of financial markets in the United States to a large extent followed the pattern in Britain, with early growth in metalworking and particularly textiles investment in the first half of the nineteenth century, funded through the savings of individuals, small partnerships and their families. In general, US developments were copies of British pioneers, though in some areas US innovations were unique, for example integrated cotton factories. Integration was enabled partly through the availability of concentrated financing from several established merchants, and led to further innovations and greater productivity (Chandler 1977, 67). However, one obvious and ultimately decisive difference was the geographical size of the US, which in 1850 had a population even smaller than that of Britain.

4.2.1 Continental Scale

The sheer area of land that makes up the United States not only provided far greater resources, but also needed fundamentally novel solutions to novel problems. From its inception in 1776, the US had largely the same institutional legal and financial infrastructure as Britain, because these were largely retained on independence. What were needed on a scale never seen before were large sums of capital to bridge the continental distances involved, particularly in the railroad business. So, while the railroads were largely the culmination of the industrial revolution in Britain, for the financial revolution in the United States they are where our story commences.

Railroad construction in the US started at very much the same time as in England. Some roads had been built to operate using horsepower even before the opening of the Manchester to Liverpool line in Britain in 1830, but after this had proved the practicability of the steam engine, large-scale building got underway in the US, as in Britain. However, the scale of the landmass in the US was in time to bring crucial differences.

Before 1850 however, the only railroad to connect one regional section of the country with another was the Western Railroad, from Worcester to Albany, connecting coastal Massachusetts with up-state New York. It had cost \$8 million to build, when Boston was still an important source for regional funding. In the 1840s over 6,000 miles of track had been built, bringing the total to 9,000 miles. The uncertainty that still existed is shown by the fact that 400 miles of canals were also built in that decade, though in the next more miles were abandoned than constructed. The 1850s saw over 21,000 miles of railroad constructed, including the almost simultaneous completion of the four major intersectional truck routes connecting east and west between 1851 and 1854; that is the Erie, the Baltimore and Ohio, the Pennsylvania and the New York Central (Chandler 1977, 82-83, 90). By this stage New York had become the dominant location where investment capital was to be found.

4.2.2 Rise of New York

In much the same way as in Britain, early railroad building in the US was financed through the savings of local farmers, merchants and manufacturers. While London had always acted as the linchpin of the English financial markets, primarily because most monetary saving was in government securities, no dominant national financial centre existed in the United States at the beginning of the nineteenth century. Significant securities markets existed in Philadelphia, Boston and New York. A combination of this decentralisation, together with a distinct antipathy in some quarters towards federal level institutions, made the establishment of an equivalent to the Bank of England problematic. Two attempts were made however, the first Bank of the United States being chartered by congress in 1791 for a twenty-year term, which was not renewed. The second was granted another twenty-year charter, but when that lapsed in 1836 no further attempts were made (Bodenhorn 2000, n168). The Second Bank was located in Philadelphia, and when it expired that city's financial role was significantly diminished. For a time Boston became the major source of capital, but by 1850 New York had established itself in its nationally dominant position.

New York's success can be attributed to two self-reinforcing features, namely its pre-eminence as a foreign trade centre and its active secondary trading markets. Both owed something to the success of the Erie Canal.

The continental expansion of the United States westward was severely restrained in its early days by limited means of transporting goods in an east-west direction. The only effective commercial route from the west was using the Mississippi River down to New Orleans, then across the Gulf of Mexico, and up the Atlantic coast. Never mind the distance involved in reaching the eastern seaboard of the United States, wheat and flour sent by this route often rotted or soured in the process, because of the heat and humidity of the southern part of the journey (Chandler 1977, 24).

Added to this, Boston and New York both had direct access to the Atlantic, and Philadelphia had indirect access through the Delaware River, and each competed strongly as transshipment points. As early as 1792 the New York State Legislature chartered two corporations to connect the Hudson River, which reached the Atlantic at New York City, with the Great Lakes, which as well as their extensive shoreline also provided a northern terminus for many of the extensive north-south river systems dissecting the continent. Despite state aid, in the form of stock purchases and loans, both suffered from inadequate planning and management, and neither could attract sufficient investment; and neither made any practical headway. Finally, the State legislature took the bit between its teeth and determined that a canal between the Hudson River and Lake Erie would be built under State auspices with a publicly appointed management board and financing provided through the sale of bonds guaranteed by the State.

The Erie Canal proved a huge success. First of all in attracting financing. There were forty-two separate issues between 1817 and 1825, raising \$7 million. Secondly the success of its operations after its completion in 1825 was sufficiently visible to lead to the construction of a whole network of canals through Pennsylvania, Maryland, Virginia and Ohio (Werner and Smith 1991, 86; Chandler 1977, 34, 45).

The launch of the Second Bank in 1816, together with the first floating of Erie Canal bonds in 1817, added a new vitality to the nascent trading in securities, and was certainly one significant factor behind the launch of the New York Stock & Exchange Board in 1817 (Werner and Smith 1991, 38). Although there is also evidence that this was a response to a similar exchange organised by the brokers in Philadelphia (Werner and Smith 1991, 28). The volume of trade attracted by the Canal brought with it the commercial paper through which the working capital supporting the trade was mediated.

With the paper came the institutions necessary to support the financial market, namely the banks. Much of the attraction of dealing in the secondary securities market was that it offered the chance, though certainly not the certainty, of a quick way to make money, with even a modest amount of initial capital. Banks found an attractive market for loans that were highly liquid, and so a profitable outlet for their excess cash balances, in 'call loans.' Speculators could leverage their securities trading by borrowing money with securities as collateral and a small down payment, but had to repay on demand.

New York quickly became the leading national market for secondary trading, although Philadelphia and Boston remained more important for primary trading in initial issues through most of the first half of the nineteenth century. However, because an active secondary market adds considerably to the attractiveness of initial issues, by the middle of the nineteenth century New York's predominance in secondary trading had led to it becoming the favoured location for all security dealing (Werner and Smith 1991, 45).

For the first time ever, a national level market for private investment finance had developed. Between 1849 and 1854 more than thirty large railroads were completed, and the sums of money involved were unprecedented. Three roads were capitalised at between \$10 and \$17 million, and the four great east-west lines at from \$17 to \$35 million.

However, domestic resources were soon too limited. Partly because of the political unrest sweeping Europe in 1848, people there found United States securities increasingly attractive. Initially government bonds were preferred, but by 1851 and 1852 first German and French savers, but followed shortly afterwards by the British, purchased private railroad securities in quantity. By 1859 investment in private railroads had exceeded \$1,100 million, including almost \$700 million in the previous ten years (Chandler 1977, 90).

4.2.3 Civil War Interlude

In 1861 the Civil War broke out, and suddenly the federal government needed considerable amounts of finance for purchasing resources, in a situation where its ability to repay was far from assured. In 1862 the Secretary of the Treasury Salmon P. Chase needed to find a new approach.

Jay Cooke had started his banking career with E.W. Clark & Co. in Philadelphia, two years after it was founded in 1837. He became a partner four years later in 1843 (Carosso 1970, 11). In 1861 he had opened his own private banking house in Philadelphia with his brother-in-law. Until that point investment opportunities had been targeted at banks and wealthy individuals. Small investors could only participate indirectly through holding deposit accounts. However, in the summer of 1861 Cooke, together with the more conservative and established Philadelphia banking house of Drexel & Co., employed extensive newspaper advertising, aggressive salesmanship and statewide agents, to place a \$3 million Pennsylvania state defence bond at par. The issue was fully subscribed in lots from \$300,000 to \$50 (Carosso 1970, 14).

Secretary Chase appointed Cooke as special agent for distributing an unsold portion of a \$500 million federal twenty-year six-percent bond issue, authorised late in February 1862. Cooke took his successful statewide approach and applied it nationwide. He appointed established brokers and private bankers to act as his agents in cities such as Boston and New York. In the less well financially served western states he appointed a variety of community leaders and local businessmen, as well as other volunteers, to act as his agents. Overall 2,500 people acted as salesmen in every state and territory (Carosso 1970, 15). Every social and economic group in the country had their patriotism and self interest called on in a nationwide publicity campaign. By mid-January 1864 the loan was closed and Cooke's organisation had sold close to \$362 million, with the treasury having disposed of the rest through more conventional means.

Cooke's fee was only one sixteenth of one percent, and his bank only really made any profit through trading this and other government securities on its own account. The combination of trading in government securities and the close relationship with the treasury that the loan issue involved led to considerable controversy, and he was not engaged as an agent again until January 1865.

Cooke set out to refine and extend his approach, including establishing 'Working Men's Savings Banks,' which in practice were night sales offices. However, the end of the war later that year brought a shift in both the government's need for capital and people's willingness to invest (Carosso 1970, 16).

4.2.4 Railroad Dynamics

The growth in railroad building following the end of the civil war continued to be accompanied by speculative dealing in securities. What was different in the United States, compared to England, was not so much the scale in itself, big as it was, but the consequence of the scale. Through traffic became a significant factor in the competitive dynamics, in a way that never occurred in England, or any other European country. It was the interaction between competing railroad firms and financial speculators that created modern corporate finance, and ultimately the modern corporate landscape.

The capital intensity of the railroad firms was unprecedented. Huge amounts of resources had to be poured into them to construct the lines and the rolling stock before they could earn a cent. Then, once operations were underway, the fixed cost of servicing the debt accrued was very high in relation to the variable cost of each ton-mile of goods shipped. These factors created a risk that competition could get out of hand and start to force prices down below a level that was able to fund the debt repayment. To minimise this risk railroad managers adopted two strategies. The first was a territorial strategy of informal alliances between roads in a limited geographical area. The

increasingly important proportion of through traffic, however, created the threat of driving prices to unsustainable levels. This was met through more formal arrangements, including national conferences to set official rates, and regional cartel agreements. Most at risk were the major east-west lines, competing, as always, for traffic between the interior and the Atlantic ports.

The combination of the fundamentally unstable strategic positions, together with the search for quick profits through security speculation, came together in 1868. Cornelius Vanderbilt had made a career as a transportation entrepreneur. First, through successfully challenging the river ferry monopoly between New Jersey and Manhattan, then later through trans-Atlantic shipping, before taking on control of the New York Central railroad in 1867. He saw the weak condition of his nearby competitor the Erie as dangerous. Should the Erie fail, its debt would be written off, or at least significantly reduced, and it would then be able to operate profitably while charging far lower prices, forcing his New York Central to reduce prices or lose traffic, either being likely to bring about its bankruptcy.

However, Vanderbilt's concern did not go unnoticed. Jay Gould, Daniel Drew and Jim Fisk had already made a name for themselves as successful speculators in the 1850s. They saw an opportunity to buy up a controlling interest in the Erie before Vanderbilt, which they duly did, albeit using many underhand tactics that they were nevertheless able to get away with. Vanderbilt withdrew from his takeover bid and, after Drew and Fisk sold out, Gould was left in control. Realising he was now in somewhat of a predicament, Gould sought to secure his position by buying up railroad companies that could feed into the Erie. This however threatened his competitors, particularly the New York Central and the Pennsylvania, which understandably were concerned that it would be problematic to come to an amicable arrangement with Gould. Vanderbilt responded by successfully buying controlling stakes in feeder roads, but kept them as essentially separate businesses. The Pennsylvania however, under the effective control of its

professional management, decided to build what would quickly become by a long way the largest single private enterprise that had existed until that point.

4.2.5 The Birth of Corporate Finance

In the five years from 1869 to 1874 the Pennsylvania Railroad went from a line with 491 miles of track to one of just under 6,000 miles. With its capitalisation at \$400 million it accounted for almost 13 percent of the total capital of US railroads. The actual amount spent was rather less, as many properties were leased rather than purchased, and only 51 percent of the stock of a firm was needed to assure control.

Nevertheless, the finance necessary was considerable, and in 1870 the company turned to Jay Cooke. The previous year Jay Cooke & Co and E.W. Clark & Co had jointly sold a \$2.5 million bond issue for the Lake Superior & Mississippi Railroad. So Jay Cooke formed the first modern underwriting syndicate in the US, and together with seven other firms underwrote a \$2 million bond for the Pennsylvania (Carosso 1970, 53). Over the course of the five years \$87 million worth of securities were sold (Chandler 1977, 155). For subsequent issues Drexel and Company became the Pennsylvania's lead banker, a role they fulfilled until 1880, when Kuhn, Leob took over.

Possibly, winning the Pennsylvania account was one factor in helping the young J Pierpont Morgan to partner with Drexel, and form a private bank in New York, when his father retired. J.S. Morgan had been running the bank he had taken over from George Peabody as an outlet for US securities in London (Chandler 1977, 155). In time, J.P. Morgan & Co. would come to dominate the investment banking industry for half a century.

The other major east-west lines, the Baltimore & Ohio and the New York Central followed similar but less ambitious system building strategies. John Work Garrett, the president of the Baltimore & Ohio from 1858, preferred a territorial strategy based on alliances, but nevertheless found it necessary to protect its traffic by building a feeder into Pittsburgh in 1866, and a new line

to Chicago in 1874. He also leased and bought controlling stakes in other lines in the late 1860s, and in 1878 gained control of the Ohio & Mississippi when it went into receivership.

The Vanderbilts, who controlled the New York Central, expanded even more slowly. Cornelius Vanderbilt even sold stock in the Wabash and other mid western roads, when his son-in-law who ran them died suddenly. William Vanderbilt who took over when his father died in 1877 made minimal purchases; the Michigan Central as part of a deal with Jay Gould, and then the Canadian Southern that connected the Michigan Central to the New York Central when that went bankrupt in 1878. The expenses involved in these and a few minority purchases to help sustain alliances, together with restoring the condition of the poorly maintained roads he took over, went a long way towards convincing Vanderbilt of the wisdom of selling off a sizable portion of his New York Central stock, which he did through the services of Drexel, Morgan and Company in 1879. The New York partner J.P. Morgan put together a syndicate to sell these shares in London. Morgan became an active member of the board, and the age when ownership and management were in the same hands for the largest companies was drawing to a close, to be replaced by the largely faceless ownership activity of the financial markets (Chandler 1977, 156-8).

4.2.6 Self Funded Consumer Business Expansion

Chandler in *The Visible Hand* describes the development of numerous industries in the United States in the final quarter of the nineteenth century. Without exception, the railroad played an important role, either as a customer as in the case of the iron and steel industry, or as a key link in the distribution chain as in the case of catalogue retailers. As Chandler points out, whether in distribution or production, what was really novel was the speed at which activity took place (1977, 235, 281). As much as any other element of the economy, this changed the nature of the financial markets, because it removed the need for the huge amounts of working capital that were tied up

in the processing and distribution of goods throughout the industrial revolution. So even though these businesses relied on internal sources for capital investment, like the earliest industrialists, they were unlike them in their credit relationships. To illustrate the scale of the changes, and drawing largely on Chandler's work, the impact of two industries that came to combine processing and distribution will be described.

The first example is that of the meat industry. Through the 1870s meat was transported east 'on the hoof'. Given that sixty percent of the animal was regarded as waste, and that the long journeys in poor conditions led to some cattle dying and many losing weight, the opportunity for profiting from safely transporting just the meat was considerable. However, the scale of the industry also created many entrenched interests. The railroads were concerned that they would lose revenue through fewer ton-miles of cargo, and existing wholesale butchers would find their whole viability put at risk. Nevertheless, one wholesale butcher from New England, named Gustavus Swift, moved to Chicago and successfully took up the challenge.

The key to Swift's success was to build an integrated enterprise. First of all he began winter shipments in 1878 while building a network of branch houses in the northeast. In the meantime he sought to improve the refrigerated car, employing a leading refrigeration engineer. This resulted in cars with a recognisably modern design being introduced in 1881, after which Swift extended his network to the rest of the country. Each branch included refrigerated storage space and a sales office with staff to sell and deliver the meat to retailers. This was also supplemented with cars that sold small lots directly in towns and villages. As well as refusing to build refrigerated cars, many railroads refused to carry them. However the competitive nature of the east-west links meant that Swift found one that would. As the market expanded, Swift expanded production, both by increasing the speed of throughput using moving 'disassembling' lines, and by building new plants in six cities along the cattle frontier.

Many existing meat wholesalers formed a national association to fight the newcomer, trying to create a significant prejudice against fresh meat killed days before and hundreds of miles away. Some existing players realised that competition was the only answer, and firms based in Chicago, Detroit, Omaha and New York succeeding in building similar networks, so that by the end of the 1880s a small number of large integrated meat-packing businesses dominated the dressed meat business (Chandler 1977, 299-301).

The second example is the sewing machine business. Machines had been produced commercially from the early 1850s, but until a legal battle over patents was settled in 1854, firms were reluctant to produce in quantity. Twenty-four firms were awarded access to the pooled patents, but within a few years three firms dominated the industry, because they had developed marketing networks in the form of agents, paid largely on commission, who were solely responsible for marketing their firm's products in a defined territory. The three firms, led by Grover and Baker, started to establish branch offices, and by 1869 Grover and Baker produced 85,000 machines a year, and the other two firms 55,000 each, accounting for seventy five percent of the industry output.

After 1860, one of the two smaller firms, I.M. Singer, started to move more aggressively in replacing regional distributors with branch offices. By 1876 the decision was taken to eliminate independent agencies altogether. Common procedures and official policies were introduced in areas such as credit sales. By this stage the structure covered not just the United States but northern and central Europe, as well as more widely spread outlets in Africa, the Near East, Latin America and the Far East. By the end of the 1880s the company had built two huge manufacturing plants, a cabinet making plant, and a foundry. In the 90s the firm obtained timberlands, and an iron mill.

All the necessary investment was financed out of current earnings provided by the cash flow from sales of machines. The cash flow was such that surplus

funds were often invested in railroad and government securities, and sometimes in other manufacturing firms.

4.2.7 Competitive Investment Banking

Meanwhile, speculator Jay Gould had not disappeared. Economic activity went into cyclical decline in 1873. One firm to feel the effects of this was the Union Pacific railroad, which together with the Central Pacific formed the first transcontinental road. Gould saw an opportunity for profitable speculation, and by 1874 held a controlling stake. Although at first he concentrated on reorganizing the road's finances and management, by 1878 he felt that he was exposed through having to depend on other roads for eastern connections. He set about a sustained campaign of system building, and by 1881 controlled 15,854 miles of road, fifteen percent of the national total. Things did not stand still for long however. He disposed of the Union Pacific by 1882. Then another recession meant that by 1884 he had been forced to dispose of most of his eastern lines, and from the mid 1880s he concentrated on putting together a system in the southwest (Chandler 1977, 159-60).

While the systems Gould built were never meant to be permanent, and were soon disposed of, the change in the competitive environment was permanent. The Vanderbilt group felt obliged to respond through purchasing other roads. One of these, known as the Nickel Plate, connecting New York, Chicago and St. Louis, had been built explicitly as a speculative venture, and indeed Vanderbilt bought it on completion as much to stop it getting into Gould's hands as for any other reason (Chandler 1977, 161). Gould's successor as chairman of Union Pacific, Charles Francis Adams Jr., was a conservative representative of Boston investors, but was soon persuaded by his managers that it was necessary to build a self sustaining system to remain competitive, which in practice meant purchasing or constructing almost 2,500 additional miles of road. To the south the Santa Fe became the largest railway system in the world, operating over 8,000 miles of track, albeit on the verge of financial

collapse. In all, the 1880s saw 75,000 miles of track built across the United States, the largest amount built in any decade in any part of the world (Chandler 1977, 163, 164, 171).

As well as providing Gould's opportunity to buy the Union Pacific, the recession of 1873 had also had another effect; the end of the first major effort by investment banks to develop a large permanent market of small investors. As at the start Jay Cooke led the way. From mid-August 1873 his depositors started to withdraw their money and, unable to stand the pressure, the house failed on September 18, triggering a panic that brought down many other firms (Carosso 1970, 25-26).

The firms that subsequently flourished, and facilitated the provision of the funds that were invested and exchanged in railroad system building, were primarily those with the capability to find investors domestically and internationally. Lee, Higginson drew on the close personal ties that existed between its partners, and directors and officers of a variety of New England financial institutions. Kidder, Peabody had similar links, but also close ties with Barings in London, for whom it acted as attorney and agent. Domestically, key customers were commercial banks, savings banks, trust companies and life insurance companies (Carosso 1970, 27).

After the beginning of the economic recovery in 1878, and up to 1893, the total of railroad bonds and stocks increased from \$4.8 billion to \$9.9 billion (Carosso 1970, 29). This represented both new building and increases in asset valuations. Alongside this, takeover activity would often involve exchange of one firm's securities for another, sometimes in the same hands but sometimes in different hands. Overall foreign investment increased between 1879 and 1900 from around \$1.4 billion to around \$3.3 billion. So although important, and additionally taking into account that much of the foreign investment went into government issues, most funds were domestically generated (Carosso 1970, 30).

A significant source of domestic funds was the commercial banking system. The figures are not wholly available, but there is evidence to show some of the changes that occurred. The other major component of the 1863 Act authorising the borrowing undertaken by Jay Cooke, after the start of the Civil War, was directed at the organisation of the banking system. Until this time banks were either chartered by States or held privately. The Act introduced federally chartered national banks, and restricted note issuing to these banks. Standard criteria for capitalisation were introduced to make the system more secure and improve confidence. The number of national banks increased rapidly, and the number of state banks declined, though they later recovered. National banks had to submit regular returns, and from these the growth in holdings of securities as bank capital can be readily observed. While in 1863 only 0.5 percent of national bank assets were in the form of securities, by 1890 they accounted for 3.6 percent, rising rapidly to 7.2 percent over the next decade, most of these held in the form of bonds, particularly railroad bonds. Between 1880 and 1910 the ratio of bonds held as secondary reserves increased from one in twenty up to one in six (James 1978, 50-1).

The critical role of external finance in railroad operations was reflected in increasingly close relationships between the railroads and the banks. Ongoing relationships between a railroad and an investment bank became the norm, and became more formalised through bank representatives sitting on railroad boards and finance committees. The banker's presence facilitated security sales for the railroad both directly and through increasing investors' confidence, while the banker could protect both their customers and their own reputations by supervising operations and discouraging unnecessary risks. Kidder, Peabody functioned as financial advisor to the Santa Fe railroad for example. During the 1870s one of the bank's partners sat on the board and chaired the finance committee. The bank had four primary roles. First, it provided new capital. Between 1885 and 1888 bonded debt increased from \$52 million to \$154 million. Second, it acted as transfer agent for the railroad's

securities. Third, it acted as bank of deposit, and fourthly it provided financial advice to management (Carosso 1970, 34).

The Santa Fe's problems started to come home to roost in late 1887. A sharp decline in earnings, among other problems, combined with the interest commitments it had entered into, brought the basic financial viability of the whole enterprise into question. In their role as financial advisors Kidder, Peabody proposed three immediate responses. First curtail expenditure wherever possible. Second, delay a new bond issue, and third reduce its planned dividend to shareholders. Unfortunately, these measures failed to solve the problems, and the bank began to become more and more involved in the operations of the railroad. Kidder, Peabody had a representative re-appointed to the finance committee, which had to authorise any expenditure over \$25,000. Improvements were adopted in the accounting system, and to raise capital for investment Kidder, Peabody agreed to sell a \$7 million six percent three-year note issue, secured by a second mortgage. At the May 1888 annual general meeting the bankers essentially took over the board of directors, appointing six new directors including two bank partners, one of whom became chairman. With effective majorities on the executive and finance committees Kidder, Peabody set about reorganising the railroad's financial structure, at the heart of which was converting more than twenty types of outstanding bond into just two. Mortgage bonds paid four percent and income bonds five percent, payments were also conditional on sufficient current earnings. The cash needed to service the debt reduced from \$11.1 million to \$7.4 million. In part because these changes made the payments vulnerable to unscrupulous manipulation, as well as the vulnerable position its difficulties put it in, the railroad was perceived to be attractive to speculators, so the bankers recommended putting stock into a ten-year voting trust controlled by them (Carosso 1970, 35-36).

Investment banks had become key players in the economy. The recession of 1893 brought severe problems across the railroad industry. Between 1894 and

1898, 40,000 miles of track, with a capitalisation of around \$2.5 billion, went into receivership, and the leading investment banks set to work reorganising and refinancing bankrupt and close to bankrupt roads (Carosso 1970, 171).

4.2.8 Corporate Financial Revolution

The capital available in the United States for private investment by the last decade of the nineteenth century was of a quite unprecedented magnitude. The investment banking intermediaries able to bring massive sums together for a single enterprise were becoming very effective.

While the enduring assets of railroads, most of all their track, was clearly visible to the naked eye, and gave investors confidence in the underlying capacity of the firm to create value, for other firms this was not the case. It was necessary to develop the techniques to ensure that as far as possible invested funds would be put to productive use. The management systems that made this commitment possible, accounting systems, had been established in parallel with developments in the capital markets. As the commitment-credit model predicts, these two developments brought about fundamental changes to the economic structure first in the United States, but then in many other countries as well. The next section describes the extent to which growth in the form predicted can be observed.

4.3 Consequences

The historical evidence is conclusive that there were indeed revolutions in commitment and credit preceding and during the 'managerial revolution' in the United States at the end of the nineteenth and beginning of the twentieth centuries. But what is the evidence regarding the nature of this managerial revolution? Did it consist of the four elements of growth predicted by the commitment-credit model, namely output growth, unit growth, financial sophistication growth and growth cycles?

4.3.1 Output Growth

The evidence that a great merger wave swept through US manufacturing industry between 1895 and 1904, substantially increasing the size of firms operating in many industry sectors, is incontrovertible. Lamoreaux identifies four mergers in 1895, six in 1897, sixteen in 1898, and in 1899, the height of activity in this period, sixty-three (Lamoreaux 1985, 1). The years 1900 through to 1902 saw twenty-one, nineteen and sixteen mergers respectively, before tailing off to five in 1903 and three in 1904. She also reports Thoreli, with slightly different criteria, as finding fifty-seven mergers from 1890 to 1893, twenty-seven from 1894 to 1897, one hundred and eighty six from 1898 to 1901, and then thirty-four from 1902 to 1903. Overall, Lamoreaux identified over one thousand eight hundred firms that disappeared within consolidations (1985, 2).

Chandler charts in some detail the evolution of what he named the managerial revolution; much of the evidence he cites has already been used to chart the development of commitment and credit mechanisms in the previous two sections. He shows that by the end of the 1880s, for the first time, a number of industrial enterprises were beginning to serve the whole of the USA. With the coming of the twentieth century many surpassed the railroads in the size, complexity and diversity of their operations (Chandler 1977, 289). Indeed, a long depression from 1873 was exacerbated to a large extent by price declines, resulting from a combination of limitations to the money supply and a rapid rise in overall output capacity (Chandler 1977, 316). During the 1880s manufacturers in a small number of industries began to consolidate. Two, namely cattle and cordage, were short lived, while six, namely petroleum, cottonseed oil, linseed oil, sugar, whisky and lead processing, resulted in the creation of a single trust that ended up dominating its industry for decades. In all the successful cases there were considerable economies of scale, and the opportunity for technical innovation in the finished product was limited (Chandler 1977, 320).

The depression of the 1890s, for Chandler, demonstrated that it was now very difficult for small, single-unit enterprises operating under a single legal roof to remain viable, because if they could maintain prices at a reasonable margin of profit then competitors would appear. In many cases, these new entrants were run by people who had previously sold out of their business to a trust, which had been created to bring a market under the control of one firm (Chandler 1977, 335). While this explanation certainly fits with the evidence, it is less clear why this had apparently only started to happen at this particular time. However, the commitment managers were able to make through audited disclosures, and the development of investment banking, was new, and explains why freestanding businesses, that had been viable, were no longer so; because previously the infrastructure necessary to make large investments, necessary to start a substantial business from scratch, had not been available to potential competitors.

As Chandler makes clear in his analysis, not all markets functioned in this way. While there were many enduringly successful mergers, there were also many that responded to the many apparent successes, but could not be sustained: only more productive merged firms survived in corporate form. A catalogue of mergers analysed by Chandler lists 156, between 1888 and 1906, which were large enough to affect the market structure of the industries in which they operated. Of these, all but eight were in manufacturing or processing. Of the 148 in manufacturing and processing, there were fewer mergers, and a higher proportion of failures, in labour intensive industries. There were none in the apparel industry, only one in furniture, three in printing and publishing, and one in lumber. In none of these did concentration of production reduce costs, or enable high volume distribution or specialist servicing to any significant extent (Chandler 1977, 337). By comparison, mergers were much more likely to be successful in areas such as food, and complex, but standardised, machinery. These were capital-intensive businesses that distributed standardised products to many customers. On the whole they were high volume, large batch or continuous process industries, or

those needing specialised marketing services. Chandler, while recognising the evidence as incomplete, believes that it is safe to say that by 1917 all the successful mergers had integrated production with distribution. Some of these successes also required a change of strategy from horizontal merger for market power, which proved unsuccessful, to a strategy of administrative centralisation for improved productivity (Chandler 1977, 338).

The conclusion Chandler draws from all this evidence is that the consolidated corporations only enjoyed continuing success where they used the resources under their control more efficiently than was possible without the combination. This was despite the fact that in many cases the motivation for the merger was in controlling competition, or the winner's personal gain from consolidation. In practice, mergers succeeded when value was created, because they enabled the control of a high volume of materials through the processes of production and distribution, from raw material to consumer. Improved information and cash flows increased the intensity with which resources were used, and so reduced unit costs. Chandler, not unreasonably, suggests that this evidence refutes the hypothesis that markets are always the most effective form of economic organisation (Chandler 1977, 338-9).

The statistical evidence from the economy as a whole also backs up the evidence set out by Chandler from within markets. Levenstein analysed US Department of Commerce figures, and found that between 1870 and 1913 total output, per capita output, and labour productivity, all grew more quickly in the US than in any other leading industrial country. Between 1890 and 1914 real output actually doubled, albeit with an increase in population. And the United States went from being a net importer to a net exporter (Levenstein 1998, 190). So ironically, while much of the initial impetus for investment banking had been to supply foreign funds used to purchase goods abroad to invest in US businesses, the US as a whole was now looking for its surplus of exports to be invested overseas, for a future return. Competitive large corporations were a benefit to the whole economy.

4.3.2 Form Elaboration

While the growth in size of firms within a market was clearly observable during the industrial revolution, it is not necessarily so obvious how the prediction of the commitment-credit model could apply to the growth of corporations with effective commitment-credit governance mechanisms. However, this pattern does appear to match up with what is referred to as the shift from U-form to M-form corporations.

The start of the 1920s saw a significant post First World War depression, after the initial euphoria accompanying the end of hostilities. Large corporations, central to whose existence was a steady flow of products to the market, found themselves having to deal with significant inventory excesses. Meat packers Armour, for example, needed to turn to outside sources of working capital, and the founding family lost control of the business as a consequence. Sears Roebuck, the retailers, avoided default on supplier payments only because its president Julius Rosenwald could draw on his family's personal fortune. Chemical and mechanical processors and manufacturers with longer and more complex supply chains found themselves in greater difficulties. Henry Ford's business survived, because he forced his dealers to buy and pay for cars they could not sell, with the threat to cancel their dealerships if they did not. As a consequence, Sears Roebuck as well as General Electric and United States Rubber developed techniques to carefully forecast future demand, and set and adjusted their flows accordingly. At Du Pont and General Motors their managers and financiers went further, creating what became known as the multidivisional structure (Chandler 1977, 456).

The multidivisional M-form corporations consisted of autonomous divisions, each of which had a well-defined market for which it integrated production and distribution, through coordinating flows from suppliers to consumers. Each of these divisions resembled the earlier U-form corporations, where overall control was centralised and activities were divided into functional departments. For the M-form corporations, a general office of top managers

with a large administrative and financial staff supervised the multifunctional divisions. The top managers evaluated the market and financial performance of the divisions, ensured that their flows were tuned to fluctuations in demand, and that they applied consistent policies in areas such as purchasing and personnel; most of all, the top managers decided between the competing opportunities for capital investment (Chandler 1977, 457). This structure came to be adopted by large enterprises through the 1920s and 30s. Holding companies such as Allied Chemical, Union Carbide and United States Steel found that it was more flexible, and so more effective, than either the holding company, sometimes called the H-form corporation, where there is no operational coordination between divisions, or the U-form centralised and functionally departmentalised structure (Chandler 1977, 463).

Alongside the development of M-form corporations was the adoption of diversification as an explicit strategy for growth. Many businesses had previously added lines that permitted them to make use of their marketing and purchasing organisations, and to exploit by-products of their processing operations. In general these had been ad hoc responses of middle managers to relatively obvious opportunities. What was new was the conscious search by top managers for new products and new markets (Chandler 1977, 473).

Du Pont was one of the first corporations to take this approach. A manufacturer of explosives, they had expanded considerably during the war, and wanted new products and markets that would make good use of existing facilities and managerial talent. Their adoption of the multidivisional structure was a response to the challenges posed by their diversification programme, alongside the general recession of 1920-21. They moved into dyes, chemicals, fibres, film and paint, but found that this overloaded the organisational capability of the centralised U-form. It was only once decentralisation had taken place that these new lines showed a profit (Chandler 1977, 473, 475).

Chandler suggests that diversification and growth strategies may well have created controversy among boards of directors, as they did in the railway companies, because of the costs of these new operations. Little evidence exists one way or the other, but capital invested in the new operations could have been returned to shareholders. However, partly at least due to more sophisticated accounting practices, the large industrials were able to maintain dividend payments while pursuing their growth strategies. Not only were previous accounting concerns with balance sheets now more correctly focused on income, but growth now required smaller amounts of capital over longer periods of time than had building or acquiring new railroad capacity. In general, diversification was funded through retained profits. So shareholders, while perhaps receiving smaller dividends, also had the expectation of substantial increases in the future as a result of continuing growth. In many cases oligopolistic positions enabled firms to survive the gruelling recession of the 1930s, because they avoided cutthroat price competition and could absorb some losses from time to time (Chandler 1977, 474).

Much of the expansion was generated internally through research organisations. While often set up originally to improve products and processes, they were additionally tasked with developing new products. In some cases considerable investment was required to achieve this, for example in chemicals or large machinery. If the products were outside the firms competence, they could be licensed to another firms, otherwise either a new division could be created, or new product could simply be added to those of an existing division, if its requirements for production and marketing were sufficiently similar to its current ones (Chandler 1977, 474-476).

In making the decisions as to which opportunities merited capital investment, large M-form corporations acted in many ways as a miniature capital market (Williamson 1988, 281). It quickly became apparent that the appropriate way to make these comparisons was through selecting those that offered the highest return on capital. Although the computational techniques and the

ultimate usefulness of a single number were clearly capable of being taken to excess, as compared with the real underlying issues, some standard techniques for organising the available data and performing calculations would be of great benefit, which naturally brings us to the third element of growth predicted by the commitment-credit model.

4.3.3 Capital Allocation

Historical evidence suggests that Du Pont was, if not the most influential, certainly one of the leading firms in the early development and application of internal capital allocation techniques. These occurred in the two decades prior to its diversification drive, described above. Indeed, Johnson and Kaplan suggest that by the end of the first world war Du Pont had developed a management accounting system that would be a model for complex modern enterprises (1987, 87). Although there were overlaps, this system had fundamentally different objectives to a financial accounting system.

Du Pont was established initially in the early nineteenth century as E. I. du Pont de Nemours and Company. Alfred Du Pont, a partner in the original company, and his two cousins Coleman and Pierre, founded the Du Pont Powder Company when they performed what in modern terms was a leveraged buyout of the original firm. Bonds were issued to the partners in the old firm paying a level of interest equivalent to the profits expected by that firm in the future. The cousins were clearly confident they could make a profit that would exceed this, and having done so would have this as their reward. The cousins bought out numerous competitors and set about creating a centrally managed enterprise. Effective capital allocation decision-making was central to their scheme (Johnson and Kaplan 1987, 66-67; Chandler 1962, 52).

The pivot of the Du Pont management system was the measurement of return on investment. While contemporary techniques enabled income to be determined, the new system required the establishment of an asset accounting system. This was inaugurated in 1903, when a complete inventory of assets

and equipment was recorded. Subsequently, new construction was charged to this account at cost, and dismantled assets credited. A comprehensive construction appropriations procedure was developed to supply the basic data (Johnson and Kaplan 1987, 69).

Each of the three major departments, namely manufacturing, sales and purchasing, provided carefully designed operating statistics. This enabled the top managers to concentrate their attention on planning future capital investment, while monitoring the performance of the divisions, but leaving decisions regarding day-to-day management with the divisions and their managers. For example, the manufacturing division provided a profit and loss sheet and a works cost report for each of the division's more than forty geographically dispersed manufacturing facilities. Only the executive committee saw the profit and loss sheet however, works managers were assessed on the basis of their works cost report, so their incentive was very much on efficient use of the facilities in place (Johnson and Kaplan 1987, 70-71).

Du Pont's policy was to finance expansion out of cash generated through operations and stock sales, avoiding external debt financing. A cash flow forecast enabled the Executive Committee to determine the maximum amount of construction to which the firm could commit itself. Accounting data provided an estimated contribution margin for each product, and the sales department provided estimates of projected sales. Figures took into account expected changes in prices and costs, and to these were added non-operational income such as land sales and income from financial securities held, so that by 1910 the Executive Committee received monthly forecasts on the firm's cash position for a year ahead (Johnson and Kaplan 1987, 69-70).

One investment decision considered by the firm was backward integration. A recession in 1907 caused a problem because of the decentralised purchasing system. As prices fell, the Vice President of Purchasing accumulated

significant stocks of raw materials. Payment required for these purchases became due just at the time that declining orders reduced revenue, thus squeezing working capital, to an extent that required considerable effort by top management to overcome the problem. Following this, a proscribed stock level was set for raw materials, and the purchasing department's remit became to buy at the lowest price, up to this level. However, this introduced the risk of supply shortages in the event of an emergency. To reduce the risk of being held to ransom by suppliers, the company began to acquire ownership of many sources of supply, although only if this met the return on investment criteria. The return required was set at fifteen percent, the normal return earned in dynamite making, which was the most profitable manufacturing activity (Johnson and Kaplan 1987, 82-3).

Another significant factor taken into account by the firm, in addition to return on investment, was their competitive position. Du Pont took the view, as had Andrew Carnegie, the founder of US Steel at the turn of the century, the first billion dollar corporation, that it was potentially counter productive to compete ruthlessly to eliminate all competition. A critical factor in sustaining the return on investment was operating expensive capital plants at as close to full capacity as possible. Smaller, and almost inevitably higher cost producers, would provide the excess capacity during market recessions. Keeping prices close to those of competitors would also ensure that there was a return on investment, so that capital as well as revenue costs were covered, and without which the business could not be sustained (Johnson and Kaplan 1987, 78).

This system was fully elaborated by F. Donaldson Brown, assistant treasurer of the firm from 1914. Although not trained as an accountant, but as an electrical engineer, his analytical training and business experience across engineering and marketing gave him the necessary perspective (Johnson and Kaplan 1987, 86-87). Between 1915 and 1918 a unique system for reporting the information for each product line and each mill was established. A series of 350 wall-sized charts were updated each month and maintained for easy

viewing in a Chart Room at the headquarters (Johnson and Kaplan 1987, 84-85).

While the focus on return on investment identified the key role of the corporation as a user of savings to create future value, it was clearly not a panacea. As noted in the previous subsection, with diversification it became too problematic to maintain the centralised structure. The aggregate picture assumed, of necessity, that changes in output or prices occurred only as predicted, and that nothing else occurred that would disrupt the normal flow of activities. Of course these assumptions were frequently violated, the over-purchasing induced crisis of 1907 being a classic example of the problems of aligning subordinate goals and overall goals in an environment subject to change. Nevertheless, the evidence would suggest that in dealing with these relatively long term investment decisions, the corporation was more successful than the market, albeit for corporations existing within a broader overall market economy.

4.3.4 Product Cycle

While the overall investment cycle for the whole economy is manifested through the business cycle, which although now subject to a degree of influence through monetary and fiscal policy, is one of the most illusive economic phenomena to pin down, in the corporation it has come to fruition in the product cycle, which like most aspects of the corporation, benefits from a far higher degree of explicit control.

Thomas Edison established what was probably the world's first active product development process, eventually having 1,100 patents to his name. In 1879 he established a research laboratory in Menlo Park, New Jersey (Treacy and Wiersema 1995, 83). By 1896 however, George Eastman, of Eastman Kodak fame, was proposing to establish an Experimental Department, with the aim of bringing out improved goods every year: depending for competitive advantage not on patent protection, but on better products

(Chandler 1977, 572). Chandler, as noted above, lists a plethora of large corporations who established substantial research laboratories to help them maintain dominance of their markets. His list includes commodity firms such as Americal Cotton Oil, National Lead, Goodrich Rubber and National Carbon, as well as technology firms such as Western Electric, Westinghouse and General Electric (Chandler 1977, 375).

However, it was not only large companies that established laboratories. The end of the nineteenth century also saw the creation of an industry of independent research laboratories, which provided contract research and development services to other companies. Between 1900 and 1940, almost 350 independent laboratories were established, employing 3,300 scientists and engineers. It is also interesting to note, however, that from 1921 to 1940, the proportion of scientific professionals employed in independent research organizations fell from 15.2 percent to 8.7 percent (Mowery and Rosenberg 1989, 79, 83).

There is evidence that anti-trust legislation played a role in encouraging firms to develop new markets, and investing in research and development to this end, rather than taking over competitors to achieve growth. A decision of the Supreme Court in 1904, brought about through the advocacy of the U.S. Justice Department, opposed horizontal mergers. Studies of both Eastman Kodak and Du Pont suggest that research laboratories were established to create new products that would enable them to grow organically. Antitrust policy did not focus on the acquisition of new technologies from other sources; so many acquisitions of this type were made (Mowery and Rosenberg 1998, 14).

The capability of large firms to develop new products as a matter of course was increasingly widely recognised as the twentieth century progressed. Whitehead in 1925 wrote, 'The greatest invention of the 19th century was the invention of the method of invention.' And Schumpeter in 1942 wrote,

'Innovation itself is being reduced to routine. Technological progress is increasingly becoming the business of teams of trained specialists' (Mowery and Rosenberg 1998, 1,2).

Perhaps the ultimate early pioneer of concept of the product cycle was the automobile industry, and the next section describes the growth of the most important source of these developments, namely General Motors. To some extent there was an interaction of the product cycle with the business cycle, but the end result was that it institutionalised innovation within a firm. Ultimately, the difference between the industrial revolution and the managerial revolution was that the industrial revolution made continuing innovation possible, while the managerial revolution made it necessary.

4.3.5 General Motors

For most of the first half of the twentieth century the paragon of business enterprises in the United States, or anywhere else, was General Motors Corporation. This serves as an excellent final case study to highlight the role of managerial commitment and credit mechanisms, and the consequences of output growth, sub-system growth, capital sophistication and investment cycles.

William C Durant formed General Motors on September 8, 1908, the same year that Henry Ford built the first Model T. Durant's first business venture had been in 1885, when as a twenty-four year old insurance salesman he paid \$50 for a patent for a two-wheeled horse drawn cart. Flint, Michigan where Durant was based was already one of the largest centres for vehicle manufacture in the US, and he and a partner contracted out the manufacturing, and concentrated on building a national sales and distribution network. Production expanded to two hundred carts a day. And by this stage the firm had its own plant for final assembly, with an annual wage bill of over three million dollars. It was apparent that, at this scale, disruptions to supplies would have serious consequences. So Durant and his partner began to set up

their own plants for supplying the necessary components. By 1900 the firm manufactured and marketed a whole range of buggies, carriages and wagons. However, that was also the first year when over four thousand automobiles were produced in the US, and a few of the many firms involved were starting to show a profit (Chandler 1962, 115-119).

By 1904 it was apparent to Durant that the rise of the automobile was both an opportunity and a threat to his existing business. The Buick Motor Company, one of the many small ventures, but one with a factory in Flint, failed that year, and Durant agreed with its creditors to take it on. His skills, experience and contacts soon transformed the business. The firm had produced 16 cars in 1903, and 31 in 1904, but by 1906 production increased to 2,295, and two years later to 8,487, to become the leading producer in the country, with Henry Ford second at 6,181 (Chandler 1962, 117-118).

Durant reasoned that the quickest and cheapest way to increase production to the degree necessary to meet the surging demand he saw developing, was through combining existing manufacturers, so this was the purpose with which he formed General Motors. Within eighteen months he had gained control of Cadillac, Oakland, six other automobile firms, three truck companies, and ten parts and accessory firms, in large part through exchanging the stock of the purchased companies for stock in General Motors. Between 1908 and 1910 sales income grew from \$29 million to \$49 million, though with much of it still from Buick. In 1910 however, there was a slight downturn in the business cycle. Sales dropped below production and Durant lacked the funds to pay his suppliers and workforce. In addition, the inventory of finished automobiles was quickly becoming obsolete as product improvements in the industry continued at a rapid pace. Durant obtained \$15 million in finance from a banking syndicate, but had to sign over control of the company to a trust controlled by the bankers for five years (Chandler 1962, 118-120).

The five years of the banker-led trust saw a degree of consolidation, but there was a lot of resistance from the managers of the constituent firms, particularly when they were profitable. One example was that attempts to institute a centralized accounting system failed. The allocation of funds for capital expenditure was officially centralized, but Chandler found little evidence of this being done in a rational or systematic way.

Durant was not idle during this time though. He acquired control of Chevrolet, another volume-production automobile firm, and financial backing through Pierre du Pont. Durant used capital made available through the security of his existing firms and his relationship with du Pont to regain control of General Motors, with du Pont as Chairman. Over the course of the war, when their explosive manufacturing business made huge profits, Du Ponts invested over \$50 million in General Motors. With the end of the war Durant expanded massively. General Motors, however, was hit by the recession towards the end of 1920 at least as badly as other corporations. To make matters worse, Durant started borrowing money in an attempt to maintain the price of General Motors' stock, which was the security he pledged against the loan. Matters came to a head towards the end of November, and Du Ponts and the bank J.P. Morgan & Co had to rescue Durant, who resigned at the end of the month (Chandler 1962, 120-129).

Piere du Pont took on the additional role of President, despite having previously gone into semi-retirement. He was however able to devolve much of the responsibility for planning the subsequent reorganisation to Alfred P Sloan Jr, a senior manager responsible for several component manufacturers within the General Motors Empire. Sloan had previously been president of an engineering firm taken over by Durant, and had been appointed by him as president of United Motors Corporation, a holding company owned in turn by the General Motors holding company. With Durant's approval Sloan had already put together a plan to integrate the whole enterprise, based to a considerable extent on his experience of bringing more central coordination

to United Motors Corporation. Durant failed to make any direct response to the proposal, but du Pont took it up enthusiastically. Sloan became a vice president, then in May 1923 took over as President, the role he maintained until 1937, when he became Chairman of the Board. A post he retained until 1956 (Chandler 1962, 129, 130, 156, 160).

Sloan was not satisfied with the level of integration before 1925 (Chandler 1962, 142). But many issues had to be resolved to make significant progress. During 1921 the managers of the major automobile divisions of Olds, Oakland, Cadillac and Chevrolet were replaced. The apparently unintentional consequence of these changes was to reduce significantly the barriers to further integration (Chandler 1962, 141). The development of uniform data definitions and procedures was recognised by du Pont and Sloan as a major challenge that needed to be addressed. Donaldson Brown was brought in from Du Ponts and set to work.

The first stage, in 1921 and 1922, was to build information procedures for controlling first the purchasing and then the production scheduling of each division, then to systematise the allocation of capital and the effective use of cash. Uniform accounting procedures were an important part of this work. Following this, the focus moved to developing data and controls to deal with future conditions (Chandler 1962, 145). For planning purposes a standard volume of eighty percent of plant capacity was used to allocate fixed costs, with adjustments being made as appropriate as the actual figures became available (Chandler 1962, 148). In 1924 Sloan thought he detected a downturn in demand again, and visits to dealers with crowded lots confirmed his suspicions. Following this, more information was collected on a regular basis, including ten-day reports from dealers and periodic reports from an independent company on new car registrations (Chandler 1962, 150).

By 1925 comprehensive plans incorporating sophisticated long-term forecasts were drawn up by each division, including factors such as industry growth,

seasonal variation, general business conditions and competitive activity (Chandler 1962, 151). In all this, each divisional manager had responsibility for his own division. Corporate standards had to be adhered to, but operational decisions were for the division. The Executive Committee, composed primarily of officers without day-to-day operational responsibility, took the more strategic decisions, but in general the divisional managers were too busy with day-to-day operations to have much time for these, so an effective division of labour had developed (Chandler 1962, 158).

By utilising the commitment and credit mechanisms available, General Motors was able to achieve extraordinary levels of growth, in the form that the commitment-credit model predicts. Output soared. Between 1924 and 1927 the corporation's share of the automobile market grew from 18.8 to 43.4 percent. Data from a similar time, between 1921 and 1925, shows that the output of the industry as a whole grew from 1.5 to 3.7 million units. In 1928, albeit the year before the crash, the corporation's profits were over a quarter of a billion dollars, which given that the industry did not exist thirty years before is quite extraordinary (Chandler 1962, 158).

To bring a degree of rational order to the internal organisation of the corporation, the approach taken by Sloan was to segment the market, in classic M-form style. Each division aimed at a particular segment of the market, largely based on price, from Chevrolet at the bottom end to Cadillac at the top. A new division of Pontiac was formed to fill the gap perceived to exist between Chevrolet and Oldsmobile, Oakland and Buick then following Oldsmobile (Chandler 1962, 143). Other manufacturing units were allocated to either the Parts or the Accessories groups, the former being those that supplied at least forty percent of their output within the corporation, rather than to resellers, as marketing and distribution were far less significant for them (Chandler 1962, 135).

By 1923 a capital allocation system similar to Du Pont's was in operation. Detailed plans needed to accompany appropriation requests, including the specific use of the funds, the amount of funds and the savings that would result. Divisional Managers could sign off smaller amounts, above that the President and the Group Executive responsible on a consultative basis for the division, as well as the divisional manager, were required to sign off. Larger still amounts needed either the Executive Committee or the Board's Finance Committee to sign off. Divisions themselves were relieved of concern for handling cash. Over a hundred banks across the country had accounts established, and the financial staff in the general office made the actual disbursements. If deposits exceeded a certain maximum the excess was automatically transferred to a selected surplus account. Interdivisional billing, although at market prices, was however settled centrally without any exchange of cash, so the corporation in many ways had its own monetary system (Chandler 1962, 147).

Finally, a key innovation that emerged during the nineteen twenties was the annual model change. Sloan describes this as having in general two one-year long phases. In the first the basic engineering and styling changes were decided on, then in the second the engineering issues necessary to bring the model into production were resolved. While each of the chassis units, the frame, engine, transmission and suspension were rarely changed all at one time, each evolved over time. These changes were coordinated between divisions and required considerable investment. Development moved from drawings to clay models to plastic models. Estimates of costs and weights were made and extensive coordination between the styling and engineering staff needed to take place, and suppliers needed planning information in good time. Prototypes were built from around eight to three months before being available for sale, and production started six weeks before (Sloan 1965, 239-246). In sum, innovation had become routine. Economic growth was the norm, not the exception.

4.3.6 Conclusions

This chapter has tested the commitment-credit model by applying its predictions to the managerial revolution in the United States from 1850 to 1950, and found that its predictions are very much confirmed in the evidence provided by secondary historical sources. The model has therefore in large part explained the economic rise of the western world, and done so in a scientific fashion, based on active observation, not the assertion of certain 'obvious facts.' Before discussing the implications of this, and to help assess the contribution of the confirmation of the model's predictions, developments in economic theory in the twentieth century need to be considered.

Chapter Five

GOVERNANCE IN ECONOMICS

Useful as the commitment-credit model has proved itself to be, in holding up to the testing which has made up the bulk of this study, it would scarcely be credible to claim that it was the first time governance has been considered as a significant economic factor. Two alternatives merit detailed consideration, and that is what this chapter aims to achieve. These will each be considered in turn, with a focus on the extent to which they take account of active observations of economic phenomena.

One approach, although not wholly successful in terms of its acceptance within the profession, possibly because it is not a single framework but a set of similar proposals, focuses on the microeconomic level at which this study has focused, and does continue to receive attention. This is the new institutional framework.

There has only been one outstandingly successful attempt to introduce a clear role for governance, although its value has been severely challenged in the last decade or two. And, while it does envisage governance operating at a different level to the observations considered here, what has subsequently been termed macroeconomics, it does address many of the issues dealt with in the commitment-credit model, such as the connection between uncertainty and investment. This is the Keynesian framework. Partly because it does deal with observations at a different level, that of aggregate national statistics rather than institutional support for interactions, the focus here is as much on why it was successful within the profession, as on directly comparing how it deals with the issues.

5.1 New Institutional Economics

New institutional economics is the best-developed approach to introducing governance at the microeconomic level and, at least on the surface, shares many features with the commitment-credit model. The biggest difference is that, while it has a central unifying idea, namely that institutions matter, the way that they matter is less well defined. However, it does map in a relatively straightforward way on to real world issues, although only in one early example in a similar way to this study, which will be considered first.

5.1.1 North and Thomas

One of the earliest and most enduring new institutional works is that of North and Thomas (1971). North has subsequently labelled the approach in this book as the efficiency view, and admits that he has now abandoned it (North 1990, 7). Nevertheless, in applying a relatively simple conceptual model to a specific historical narrative, namely the prelude to the industrial revolution, of all the works considered in this chapter, it is the closest in approach to that taken in this study. The record inside a recent copy of the book of five reprintings in the 1990s suggests, in addition, that many still find a use for it.

North and Thomas focus on property rights, and specifically on the incentives for people to benefit from their own innovations (North and Thomas 1973, 1). They claim that their 'comprehensive analytical framework' is 'consistent with and complementary to standard neo-classical economic theory.' It is however presented as no more than 'an interpretive study,' primarily to provide a contrast with both traditional and new economic history, where the former is much more detailed and exhaustive while the latter works by way of precise empirical statistical tests (North and Thomas 1973, vii).

The central thesis is set out right at the start of the book: 'Efficient economic organization is the key to growth; the development of an efficient economic

organization in Western Europe accounts for the rise of the West' (North and Thomas 1973, 1). And the key institutional mechanism necessary to bring this about is the one that brings the private rate of return close to the social rate of return, namely property rights.

A list of factors are identified as leading to growth, namely technology, human capital, reduction of information costs, economies of scale and population expansion; although growth in effect means per capita growth so population is not as straightforward a concept as it may at first appear. But they argue that there must be a gap in the theory because 'we are left wondering: if all that is required for economic growth is investment and innovation, why have some societies missed this desirable outcome?' North and Thomas suggest that these factors are what constitute growth, and none of them provide a causal explanation. The missing link is that 'individuals must be lured by incentives to undertake the socially desirable activities' (North and Thomas 1973, 2). But, 'if the private costs exceed the private benefits, individuals ordinarily will not be willing to undertake the activity even though it is socially profitable' (North and Thomas 1973, 3).

The underlying pressure for changes in property rights, according to North and Thomas, is a result of an increasing scarcity of resources, which in the case of Western Europe was a result of population growth (North and Thomas 1973, 19). Together with this was a difference in climatic conditions and density of population, which brought an opportunity for specialisation, particularly in agricultural products, which vary in their suitability by geographical area. The rise of the market enabled this specialisation to occur, and gains from trade reinforced the success of the process (North and Thomas 1973, 26).

An integral element of the market was money, which had not been a necessary or widely used mechanism in the feudal economy, when a peasant exchanged some of his labour for the security his lord provided. Money began

to replace the more varied but less flexible feudal exchanges (North and Thomas 1973, 28).

Despite these developments, North and Thomas argue that a tension built up during the thirteenth century, because of the misalignment between individual and social incentives. Individuals continued to perceive a benefit from large families, but landowners saw the disruption caused by the improvements necessary to produce more food as insufficient to pursue them. Although significant beneficial changes were made in the area of trade and industry, the overwhelming dominance of the agricultural sector more than cancelled them out (North and Thomas 1973, 70).

The tension between population growth and conservative landowners reached a crisis point in the fourteenth century, when famine, disease and war brought about significant reductions in the population (North and Thomas 1973, 71). The shift in the balance of power from lords to labourers finally brought about the end of the feudal system of obligations, leaving only the monetary payment of rent from the occupier to the owner (North and Thomas 1973, 80).

Population growth started to return in what is now termed the early modern period, and assisted the expansion of the market. North and Thomas assert that improving the efficiency of exchange was the source of the greatest economic gains. They identify two sources of costs, the reduction of which provided efficiencies, namely production costs and transfer costs. Another name for transfer costs are transaction costs, which are made up of search costs, negotiation costs and enforcement costs. Unlike the production of agricultural and industrial goods, the transaction sector is subject to economies of scale because they have a significant fixed component, so the cost of this fixed component decreases as a proportion of total cost as volume increases. So, where declining productivity in agriculture does not outweigh the benefits, it is possible for per capita income to increase with

population growth in the absence of technological change (North and Thomas 1973, 93).

It was differences in the type of property rights created by the different emerging states of Western Europe between 1500 and 1700 that caused the differences in economic performance (North and Thomas 1973, 97). Population growth appears to have been the norm during the sixteenth century, but the seventeenth century saw distinct divergences along national lines. Germany and Spain saw the population cycle go into decline again, with famine, disease and war returning on a significant scale. In France and Italy population remained stable, while in the Dutch Republic and England growth was sustained; although in spite of, rather than avoiding, disease and war (North and Thomas 1973, 105).

In France a strong central government resorted to increasingly rigid controls on economic activity, to enable the government to extract the maximum tax revenue. Ironically, to keep pace with England and Holland's rising tax revenues from rising income, it resorted to increasing extraction, and so further stifled economic improvement. North and Thomas identify the major problem as the restriction of product markets, on the basis that property rights were protected and factor markets, with the exception of capital, were able to develop (North and Thomas 1973, 127).

The situation in Spain was in many ways worse. It needed to sustain an extensive empire, including large tracts of the Americas and the Netherlands. Extraction from these provinces sustained their defence but, particularly when silver and gold extraction from the Americas suddenly reduced in the early seventeenth century, there was little in the way of economic infrastructure at home, and what there was became subject to desperate government interference, bringing considerable reductions in the little economic activity that already existed (North and Thomas 1973, 131).

North and Thomas cite the United Provinces of the Netherlands as the first area of Western Europe to experience both population growth and per capita income growth (North and Thomas 1973, 132). Political developments created the opportunity for markets to flourish. These included the successful rebellion of the northern provinces from Spanish rule, and their organisation as a republic. Costs were reduced in all three of the transaction type areas, namely search, negotiation and enforcement. Working capital markets developed in the sixteenth century. Lending to governments also developed, though due to the fact that the Spanish crown was a large borrower, but regularly defaulted, development here was not so rapid; but it was well underway later in the seventeenth century (North and Thomas 1973, 141). Agriculture became increasingly capital intensive and specialised. For example, large areas were reclaimed from the sea (North and Thomas 1973, 143). Trade was the primary economic activity, with Amsterdam becoming the primary market in Europe. But industry also developed, particularly textiles, with the guild-protected towns concentrating on the luxury end of the market, and the unrestricted countryside serving the value-for-money end (North and Thomas 1973, 144).

However, it was England where economic growth really took off. The clear distinction North and Thomas perceived between England and the Netherlands was the institution of a workable form of intellectual property, through the Statute of Monopolies of 1624. This proscribed royal grants of monopolies, except in the case of genuine innovations (North and Thomas 1973, 148). Over the seventeenth century most of the post-medieval European commercial and institutional innovations were brought to England from the Netherlands, where on the whole they had been copied from the Italians (North and Thomas 1973, 155).

From the detailed summary just set out, it is clear that there are significant similarities in North and Thomas' book to the approach taken here in testing the commitment-credit model. It is by far the closest equivalent in the

literature. Nevertheless, there are important differences and problems with their approach, which in fact has led North to move away from the approach. The alternative he has found to his 'efficiency' approach, namely transaction cost economics, will be set out before discussing new institutional approaches as a whole.

5.1.2 Coase

North and Thomas' use of the concept of transaction costs, though not specifically cited as such, is a clear reference to the work of Coase, and in particular to his 1937 paper, which is an early attempt to explain the increasing prevalence of large firms, when the neoclassical model would suggest that smaller units operating in a market environment would always be the most efficient economic form.

Coase introduced the concept of 'transaction costs', the term he invented to describe the cost of using the price mechanism (Coase 1937, 38). Price discovery and negotiation costs are two elements he identifies, as well as a willingness to pay more to reduce uncertainty. These costs do not exist within the firm, or at least not to the same extent. There may be additional incentives for firms to form because of tax effects, for example through a sales tax that was levied on market transactions, but not on transactions within a firm, but these are unlikely to lead to the creation of firms by themselves (Coase 1937, 41).

While transaction costs limit the price mechanism, diminishing returns to management limit the size to which firms can grow. Three factors contribute to determining the equilibrium point at which market and management costs balance, and so the size to which firms will grow. First, the lower the cost of organising, through such things as the cost of telecommunications technologies that facilitate internal communication, the larger firms will be. Second, the less likely the entrepreneur is to make mistakes, the larger the

firm will grow. Thirdly, the larger the benefit from bulk purchasing the larger firms will grow (Coase 1937, 45).

Two alternative explanations of the firm are considered by Coase, but rejected as being in theory capable of being dealt with efficiently through market mechanisms. First is the division of labour. But this fails as an alternative explanation because the whole essence of the market mechanism is the decentralised coordination of different activities (Coase 1937, 47). The second possible explanation is the existence of uncertainty, but in principle this can be dealt with through contracts between market participants using the price mechanism, so this also fails to explain the existence of the firm (Coase 1937, 51).

For Coase, two dimensions determine the quality of a theory, namely its tractability and its correspondence with the real world (Coase 1937, 33). The correspondence of his model with the legal definition of employer and employee is presented as evidence of its correspondence with the real world (Coase 1937, 54). The precise definition of the costs of a transaction in determining the size of the firm enables the marginal product of the entrepreneur to be defined, and so provides a clear integration with what is now called the neoclassical model, so making the concept tractable (Coase 1937, 55, 34).

5.1.3 Williamson

The work of Williamson draws firmly on that of Coase (Williamson 1988, 3), although he makes it clear that transaction cost economics is an alternative to the neoclassical model (Williamson 1988, 2). He characterises this as the idea of the firm as a production function being supplanted by the idea of the firm as a governance structure (Williamson 1988, 16). It is clear that governance structures generalise the concept of property rights.

Three aspects of economic transactions, or contracts as Williamson most often chooses to label them, are identified; namely bounded rationality,

opportunism and asset specificity. This characterisation leads naturally to a distinction between four approaches to economic analysis. Three take into account two of these, but not a third, while only the governance approach is able to deal effectively with all three aspects.

Bounded rationality is the characteristic most often associated with the work of Herbert Simon. It is the idea that people act in economically rational ways, but that there is a limit in their capacity to deal with all the information that may be available, so that although they aim to act rationally they do not always succeed in doing so.

Opportunism is defined as self-interest seeking with guile. The idea being that what people say they will do, and what they choose to do, may not be the same thing. This is not to say that people always act like this, but that some people do sometimes.

Asset specificity recognises that in order for a transaction to take place, investments often need to have been made previously, which will only return their cost over several transactions of that type, and cannot be utilised for other transactions. This is the most critical dimension of transaction cost economics (Williamson 1988, 30).

The three approaches to economic analysis that fail to deal effectively with one of the three characteristics are planning, promise and competition. Planning deals with opportunism and asset specificity, but assumes perfect rationality. Promise can cope with bounded rationality and asset specificity, but assumes people will carry out their promises. Competition copes with bounded rationality and opportunism, but because all transactions are instantaneous exchanges, cannot cope with asset specificity.

Governance, by combining aspects of all three can deal with all three characteristics of economic transactions. While bounded rationality and opportunism are generally well-understood phenomena, and can be handled

within the neoclassical market framework, the concept of asset specificity may benefit from a little more explanation. As in many other cases, the concept can be found in the work of Marshall, but was not developed, and indeed disappeared from sight as the neoclassical model was developed (Williamson 1988, 53).

Asset specificity requires a dynamic treatment, because it only makes sense in an intertemporal context. The central issue is the trade-off between the cost savings of a specialised technology and the hazards that arise from the irreversibility of the investment. The governance approach is able to systematically vary the nature of the trade-off. Williamson points out that, although perhaps superficially the same, there is a distinction between asset specificity and the accounting concepts of fixed and variable costs. Asset specificity is a consequence of redeployability rather than ease of allocability, which is the basis of the accounting distinction. In practice, in much the same way as the idea of fixed costs, specificity is not an either/or attribute but a matter of degree (Williamson 1988, 54). This clearly has consequences for its observability.

Asset specificity leads to what Williamson terms the 'fundamental transformation.' This is the point at which the producer is committed to providing the specific asset. While the dynamics of an auction where there are several bidders is well understood without asset specificity, once it is introduced the dynamics change, at least given bounded rationality and opportunism. After the fundamental transformation the identities of the parties matter, they are no longer faceless market participants (Williamson 1988, 62). The frequency of the type of transaction, together with the degree of specificity, then implies the efficient governance structure. In particular, ownership, observed as vertical integration, becomes efficient for recurrent idiosyncratic investments in specific assets. Mixed investments, where there is some standardisation from which economies of scale result, as well as some degree of idiosyncratic specialised investment being required, benefit either

from bilateral contracts enforced through reputation effects, or need a third party to arbitrate. Reputation can work where the transaction is recurrent, but arbitration is necessary where transactions are only occasional (Williamson 1988, 79).

Within transaction cost economics Williamson makes a distinction between two branches, namely the governance branch and the measurement branch. Both emphasises are only necessary when the three issues of bounded rationality, opportunism and asset specificity are present. Although written in the mid 1980s, and still the leading source, in regard to the issue of observability, it is clearly problematic that he states, 'the relevant dimensions for ascertaining where the measurement difficulties reside remain somewhat obscure' (Williamson 1988, 81).

Williamson argues that 'transaction cost economizing is the previously neglected but key concept for understanding organizational innovation in general and vertical integration in particular' (Williamson 1988, 129). Organisational innovation in this context refers to innovations in organisational form and function. Nevertheless, the explanation Williamson provides is through comparing the before and after states of organisation, and indeed why, contrary to neoclassical doctrine, large organisations exist at all. It does not address the process by which more general innovation is possible.

Williamson is clearly aware that there are also conceptual limitations to the transaction cost approach, though this is not made particularly explicit. For example, it is clear from the context that transaction cost economics is included in the statement 'The economic factors that lie behind the size, shape, and performance of the modern corporation, however, are poorly understood' (Williamson 1988, 295). Transaction cost economics, as Williamson observes, does take the corporation seriously, largely as a result of asset specificity. And to some extent it can explain the benefits of the M-form multiple unit firm, compared with the H-form holding company or the simple

hierarchical U-form, through the factors of bounded rationality, which limits the U-form, and opportunism, which limits the H-form. Nevertheless, there is no explanation, never mind prediction, of why the form would appear at one point in history.

5.1.4 North: the Sequel

For Douglass North 'history matters,' and his more recent work pays close attention to transaction costs, although he believes that gains from trade, as introduced by Adam Smith, still form the primary source of economic development (North 1990, vii). However, the primary argument of his most recent book is that even the combination of transaction and production costs is not sufficient to explain economic phenomena. He attributes to Coase the observation that, when it is costly to transact, institutions matter (North 1990, 12). However, he does believe that it is important to recognise that costs are only one side of the equation: benefits the other. It is the institutional structure of a society and economy that shapes the incentives which determine performance (North 1990, 135).

To a large extent, institutions are a response to limited rationality. North highlights two particular limitations, namely motivation and understanding the environment (North 1990, 20). Motivation is an issue because of the evidence that people are not always as directly self-seeking as conventional economic analysis assumes (North 1990, 22). Understanding the environment is the more commonly cited limitation on rationality, namely that without complete information, perceptions of reality vary between people, so that not only can an individual not make optimal choices, but also the choices that other people make, and that have an impact on others, are for this reason fundamentally unpredictable (North 1990, 23).

Institutions are introduced as a cost effective way of dealing with limited rationality, and particularly the limited environmental understanding. They provide standardised ways of dealing with common situations that avoid the

need for negotiation, or analysis, on how best to proceed. And they come in two varieties, namely informal and formal (North 1990, 36). Informal constraints take three forms. First they can be extensions and elaborations of formal rules. Second they are socially accepted norms, and thirdly they can be internally enforced behavioural standards. In many cases they are self enforcing, because the costs to an individual of not following them exceed those of conformity. In other cases they are more problematic to enforce, but can still be effective if reputation is a significant factor in interactions (North 1990, 41).

The distinction between informal and formal constraints is in practice only one of degree, and has come about because of the growth of complex societies with extensive specialisation and division of labour (North 1990, 46). Particular rules are not necessarily efficient, but generally represent influential groups' perceived interests. North distinguishes three levels, namely political rules, property rights and contracts (North 1990, 48). When constraints are introduced, some attention to compliance and enforcement costs is generally given, but enforcement is often not costless, and its lack is the major block on economic development (North 1990, 54).

Institutions form a complex web that through their interrelations provide a stable background for conducting economic transactions. Nevertheless, continual incremental changes are taking place, which over time can bring about significant changes. The higher the costs imposed by the institutional constraints, the more informal institutions will be used, and it is possible that if costs are high enough, then no economic exchange will take place. Transaction costs are the most visible manifestation of institutional costs, although while some transaction costs are clearly measurable, others such as search costs and delays are much more difficult to measure (North 1990, 68).

The transaction costs that result from the network of formal and informal institutions are the underlying reason, as Coase asserted, for the existence of

the firm. Individuals create firms to maximise their wealth, as defined by the institutional structure of their society. In pursuing this objective organisations will incrementally alter the institutions through which they operate (North 1990, 73). Organisations benefit from two specific factors, namely tacit knowledge and repeated team play. Tacit knowledge, unlike communicated knowledge, can only be developed through practical experience. And organisations benefit in competitive situations from developing routines and modes of action that enable more effective coordination in the face of limited rationality (North 1990, 74).

Entrepreneurs are the agents of economic change. Changes boil down to two possibilities, namely changes in relative process or changes in consumer tastes. These can both bring about changes in bargaining power, and it is this as much as anything else that brings about changes in institutions, sometimes even formal political institutions (North 1990, 84). Although radical discontinuous change can occur, overwhelmingly institutional change is incremental (North 1990, 89).

One of the major motivations behind North's recent work is to develop a theory that, unlike his previous one, is able to explain the enduring differences in economic performance levels between countries. He believes that it is the nature of institutional change that provides the explanation. Two factors dominate the paths of institutional change. The first is path dependency, which is the situation where changes happen in a way that is contingent on what has gone before, not simply on what the future possibilities are. The second is the significant transaction costs present in imperfect markets (North 1990, 95). One example of this is the adoption of constitutions very similar to that of the United States in many countries in Latin America in the nineteenth century. However, because the enforcement mechanisms, and subjective models and norms of actors were different, the outcomes of the adoptions were very different from the United States (North 1990, 101).

Underlying the success of the western world, North identifies three organisational innovations. The first were those that increased the mobility of capital, including bills of exchange and accounting techniques for monitoring agents' behaviour. The second were those that lowered the costs of information, such as printed price-sheets and technical commercial manuals. Thirdly, there were those that spread risk, such as insurance (North 1990, 125). He also perpetuates the myth of a distinctive legal doctrine for the law merchant as playing an important role in enforcement, through relying on merchants to determine accepted practice among themselves (North 1990, 128).

North argues that his contribution is largely in providing an alternative to cliometric economic history, which applies the neoclassical model to historical data. As he says: 'Although the systematic application of price theory to economic history was a major contribution, neoclassical theory is concerned with the allocation of resources at a moment of time, a devastatingly limiting feature to historians whose central question is to account for change over time' (North 1990, 131). Indeed the only major body of economic theory to take institutions seriously is that of Karl Marx, but that is problematic not least because it requires a fundamental change in human behaviour to bring about a utopian ending (North 1990, 132). North concludes with the observation that his own theory is far from comprehensive, at least because the serious study of institutions is only beginning (North 1990, 140).

5.1.5 New Institutional Economics and Observability

It is clear from the analysis just presented that many elements of the commitment-credit model and its applications in previous chapters have been proposed previously, although not in the same specific form. The real claim to novelty of the model, and its improvement over the works considered in this section, is not in giving different weights to previously recognised factors, but in relation to the application of abstract models to making observable predictions.

North and Thomas take an approach that looks very promising, and is in many ways similar to the overall approach taken here. Their conceptual framework, combining property rights and efficiency, is straightforwardly defined and tested. To some extent, because the work is not set out as a test of predictions, it is more difficult to assess its success. For example, the attribution of England's success as due to intellectual property rights is not couched as a prediction, but as an explanation, although it is perhaps possible that it could have been predicted from the framework. The failure that North himself attributes to the work is that property rights are considered in terms of their costs and not their incentives, so although the framework survives testing, the logical model is under-specified.

In many ways the problem of the property rights approach as a whole is that identified by Atiyah (1979, 138). In offering legal recourse to enforcing contract rights, there is an implication that society can enable reasonable expectations to override property rights. One of the reasons for the power of the commitment-credit model is that it recognises the central role of contracting over property rights in economic growth and development.

Although Coase's work has been very influential, as an observable theory it is enormously problematic. There is a clearly and concisely defined conceptual framework, but, at least in Coase's work itself, no testable prediction. A crucial reason for this is probably the problematic nature of measuring costs, as section 4.1 highlighted in describing the evolution of managerial accountability. Despite Coase's contention that it is a complement, rather than a substitute, for the neoclassical model, transaction costs do perhaps explain why if the neoclassical model were testable it would quickly be falsified. As North wrote in a work between the two considered in detail in this chapter, 'a neoclassical world would be a jungle and no society would be viable' (North 1981, 11).

Williamson, although ostensibly following the transaction cost tradition, in fact makes a considerable jump, by recognising the role of incentives as well as costs, and in the process recognising the incompatibility with the neoclassical model that this implies. The introduction of the notion of asset specificity clearly recognises that investment or innovation goes hand in hand with uncertainty and cost, which is also the primary focus of the commitment-credit model.

Having said this, the framework Williamson proposes is not really clearly defined. His objective, it is clear from the remainder of the book, is policy recommendation, rather than any form of active observation. He explicitly recognises the weakness of observability, or measurement as he describes it, in his framework, but clearly sees this as a secondary issue. His conceptual framework does seem to be at a similar level of complexity to the commitment-credit model presented here. However, his treatment of benefits is largely limited to negative costs, and the role governance plays in shaping these.

North's most recent approach, utilising a conceptual framework much closer to Williamson's, but with an explicit focus on institutional logic, should be very promising. Unfortunately, it makes no real attempt to relate the framework to specific situations, and it is difficult to see any concise conceptual core. While North believes the weakness of the work to be a lack of comprehensiveness, it is in fact a lack of specificity. While his early work was essentially under-specified, he has now succumbed to the opposite temptation, trying to incorporate too many factors, all of which no doubt play some role, but in the process fails to develop a coherent self-supporting logical framework.

Like the commitment-credit model, North's work does have the considerable benefit however, over much of contemporary microeconomics, which focuses on the consequences of imperfect information for small numbers of

actors, that it is motivated to a large extent through a desire to explain long-term economic development. That is, it focuses on the relations between individual economic actors, or at least on the effects of many interactions of a particular form, and the economy as a whole. So in many ways this thesis is firmly embedded in the tradition of new institutional economics, and in particular solves the main problems identified here with North's most recent work.

While contracts and other commitment devices naturally and conventionally suggest a microeconomic approach, growth and financial markets are more conventionally considered from a macroeconomic, or whole economy perspective. So, while the commitment-credit model clearly deals with the microeconomic domain, it is to the macroeconomic domain we now turn, to enable an assessment of the contribution of the commitment-credit model to be made that takes some account of work in this area.

5.2 Keynesian Economics

The most successful twentieth century attempt to incorporate governance in some form directly into economics came from John Maynard Keynes' work in the 1930s. This was clearly focused on a different level to the commitment-credit model, focusing on the role of national governments in using public spending to stimulate economic activity in certain situations. In the process, Keynes largely created the distinction between microeconomics and macroeconomics, because while the neoclassical framework continues to dominate the teaching of economic analysis at the level of the firm and the consumer, it has largely been abandoned, or at least bypassed, at the level of the economy as a whole. Despite the differences between Keynes' focus on aggregate measures for an economy as a whole, and the commitment-credit model's focus on the systemic consequences of the institutional factors experienced by individual people and firms, several of the concepts

introduced by Keynes do bear a similarity to those in the commitment-credit model.

But as well as the inherent interest in making these conceptual comparisons, a consideration of the reasons for the success of the Keynesian model will provide a comparator against which some assessment can be made of the potential contribution of the commitment-credit model to the economics discipline, which is the subject of the next chapter. Given the massive amount of work extant on Keynesian economics, even a survey would be problematic, so it is not attempted. Instead the focus here is much more on preparing for the assessment of the commitment-credit model and its contribution, by focusing more on what might be called the sociology of the Keynesian revolution. That is, examining what it actually entailed at the level of the economics profession, and to what extent it is possible to explain why it took place.

5.2.1 Theoretical Background to Keynes

According to Skidelsky (1992, 418), prior to Keynes' publication of his *The General Theory of Employment, Interest and Money*, the basic framework of economic theory, with no distinction between micro and macro, was that set out by Alfred Marshall in 1890 in his *Principles of Economics*. This was particularly so in Keynes' base at Cambridge University, which was also Marshall's home. Marshall's contribution, and in many ways the reason for his subsequent role, was that he synthesised existing, and apparently conflicting, alternative theoretical conceptions. Most important was his reconciliation of the classical theory of value with utility theory, and in particular the marginal utility formulation (Skidelsky 1983, 43). Also of importance was his reconciliation of Jevonian and Ricardian value theories, though this reconciliation was by pulling them apart, rather than bringing them together, which is possibly why Marshall's insight in this case seems less powerful than the classical and utility case (Skidelsky 1992, 419).

The classical theory of value, applied in almost identical ways by economists as diverse in other aspects of their work as Adam Smith and Karl Marx, holds that the value of a good expressed as its price is fundamentally determined by the cost of its component elements. Smith, for example, recognised that from time to time the price in the market may diverge from the production determined 'natural price' because of fluctuations in demand, but held that with competition the natural price would coincide with the market price (Blaug 1997, 38-43). Marx argued that concentration of ownership of capital prevented workers receiving the benefits of all the labour they put into a product, this being consumed by the capitalist class on non-essential, that is luxury, goods (Blaug 1997, 221-227). This theory of value, made explicit to a greater or lesser extent, was that employed by many writers from the earliest times.

Then, in the early 1870s, three men introduced the principle of diminishing marginal utility as the fundamental building block of a formal economic model. William Stanley Jevons in Manchester, Carl Menger in Vienna and Leon Walras in Lausanne came to this discovery by very different routes. The success of this development was a consequence of its relatively straightforward mathematical formalisation, and the amenability of the formalisation to analysis using the differential calculus (Blaug 1997, 277-8). Utility itself was a concept developed by Jeremy Betham earlier in the century, as a way of formalising the process of deciding between conflicting choices faced by society.

While marginal utility theory addressed the issue of demand in a precise way, it had nothing to say on supply. The leading approach to the supply side was that of David Ricardo, who at the beginning of the nineteenth century had brought together what was in effect the first rigorous formulation of economic logic, largely to address the most contentious political issue of his day, the British Corn Laws. For Ricardo there were only three factors to consider, namely labour, land and capital. Land was given, and capital and

labour were potentially augmentable. Ricardo's insight was that less and less productive land would have capital and labour invested into it, up to the point where the value of the crops harvested equalled the value of the labour and capital input. For all the land more productive than this the landowner would be able to charge rent on the difference between value of inputs and output. In this way the role of the margin, at least for the supply side, was introduced into economics.

Marshall's achievement was to reconcile marginal utility and Ricardian supply, along Smithian lines, by proposing concurrent roles for marginal utility, in regulating demand, and for Ricardian supply. Both could operate in parallel, in such a way that equilibrium exists where the two effects balance. A key conceptual device in this reconciliation was the introduction of distinct time periods. In particular, a 'market period' in which supply is fixed, a 'short run' in which productive capacity is fixed, but quantities can be varied, and the 'long run' in which productive capacity can be augmented. Blaug also identifies a fourth period in Marshall's thinking, where production techniques can be altered, though this was not made explicit. However, by largely ignoring this possibility, Marshall's theory became one of production and consumption, and in the process the primary concerns of the classical economists, which Marshall also shared to some extent, became obscured (Blaug 1997, 278, 354). It is Marshall's formulation that continues to provide the foundation for microeconomics.

Macroeconomics did exist to some extent before Keynes, or at least macro type questions were addressed. There was however little in the way of a coherent theory. The questions addressed were rather obscure, such as the degree of 'roundaboutness' in the economy, the level of hoarding and over-indebtedness (Blaug 1997, 647). No connection was perceived between monetary policy and overall economic performance. The reason underlying this seems to have been the implicit assumption that the price mechanism was the most important economic phenomena, and prices would change far more

quickly than quantities. In many ways it is a great irony that Smith's work was contemporary with the industrial revolution, and perhaps most symbolically with Watt's engine, and so was made obsolete while it was being written; then Marshall's *Principles* were first published at the beginning of the 1890s, when the mechanisms were almost in place that led to the great merger wave in the United States, creating a vastly different economy to that dealt with by his theory.

Keynes initial focus as an economist was monetary theory. Prior to his time, this was on the whole treated as a short run problem of disequilibrium. The issue was how monetary policy could stabilise the price level when outside fluctuations changed the equilibrium point (Blaug 1997, 616). Marshall in his *Principles* did not deal with this, and his approach was only published in book form in 1923, when Marshall was over 80. However, it was taught through 'oral tradition,' augmented by published records of Marshall's evidence to government committees on monetary policy (Skidelsky 1992, 418). Wicksell, a contemporary of Marshall's, had addressed the issue of interest rates, but Marshall largely ignored them, focusing instead simply on the level of cash balance that each individual endeavoured to maintain to facilitate regular transactions (Blaug 1997, 618). However, before setting out Keynes' innovations, it will be worthwhile to describe the broader economic environment in which he lived.

5.2.2 Historical Background to Keynes

The nineteen twenties saw very mixed economic experiences in Britain and the United States. In Britain, unemployment remained around ten percent for the whole decade, and by 1925 industrial production was fourteen percent below its level in 1913. In the United States production had in this time risen forty-eight percent. Until the crash of 1929 prosperity was widespread, helped by new goods such as electricity and automobiles being widely consumed (Backhouse 2002, 214). This all changed in the 1930s.

Skidelsky describes Keynes as: 'an applied economist who turned to inventing theory because the theory he inherited could not properly explain what was happening' (Skidelsky 1992, 405). While this description fits his investigative style, it is not wholly consistent with the fact that his first major research project dealt with the abstract subject of probability theory. Although that was apparently inspired by what could be described as a practically driven problem, namely his determination to enlarge the scope of individual judgement and narrow the concept of custom as the 'guide to life' (Skidelsky 1992, 56). Perhaps it is better to say that he was a theoretician who was most inspired not by other's theoretical work but by practical problems. Economics has never been short of significant practical problems, and in many ways Keynes' economic theorising was inspired by the two most widely recognised problems of his lifetime, first of all the business cycle and then by unemployment (Skidelsky 1992, 318).

Up until the 1930s, unemployment was not seen as a problem in itself, but one of adjustment to the constant ups and downs of the business cycle, particularly from the 1890s to the 1920s. The problem was seen to be in the fluctuations, particularly as this was influenced by monetary policy. The problem was seen by many as one of prices, and particularly the price of labour; in other words wages. However, by the 1930s the problem was starting to be perceived as one of the normal level of unemployment (Skidelsky 1992, 318). In the US, industrial production, following the crash, quickly fell to half its level of 1929 and unemployment rose to over 25 percent, and by 1936 had still not fallen below 15 percent. Other countries had a similar experience. In 1933 unemployment was at 26 percent in Germany, 27 percent in the Netherlands, 24 percent in Sweden, 33 percent in Norway and 21 percent in Britain (Backhouse 2002, 215).

Perhaps of most direct relevance among the events preceding Keynes' development of his *General Theory*, was the first tentative systematic development of national economic statistics. Various estimates had been

made from time to time, based on a variety of guesses and inferences from sources such as tax and census data. But only in the 1920s were the first systematic studies undertaken, first of all by the National Bureau of Economic Research in the United States. The US government became directly involved with the onset of the depression and its social and political consequences, and in June 1932 a Senate resolution committed the Bureau of Foreign and Domestic Commerce (BFDC) to submit estimates of national income for 1929, 1930 and 1931 (Backhouse 2002, 241). At this early stage, there was little consensus on what data actually constituted national income. The BFDC produced two figures, one measuring the net product of the whole economy, and another of payments to those who produced the net product. A third estimate was produced by the Brookings Institute of the sum of final products passed on to consumers and businesses. This latter figure was the first to be described as Gross National Product (Backhouse 2002, 242).

In the United Kingdom, work got underway more slowly. Keynes was however able to draw in particular on the work of Colin Clark. Figures were felt to be very politically sensitive. For example, the government denied the existence of the national income figures it calculated in 1929, because they estimated wages as being lower than other existing estimates. This even went to the extent that, according to Backhouse, despite Clark being appointed to the secretariat of the Economic Advisory Council in 1930, the Treasury refused to buy him an adding machine. Nevertheless, in 1932 Clark was able to produce figures for UK GNP, broken down into the categories of consumption, investment and government spending (Backhouse 2002, 243).

5.2.3 Keynes' Theory

Keynes' first major theoretical contribution to economics, *A Treatise on Money*, was published in October 1930. Skidelsky characterises it as an attempt to theorise about economic instability, though in a pre-depression worldview, and drawing in particular on Britain's experience in the 1920s. In line with the thinking of the time, and the neoclassical view, the central problem was seen

as adjusting to price changes, particularly when they are in a downward direction. Although agreeing with the diagnosis of the problem, Keynes broke with the neoclassical approach of analysing everything in terms of a static equilibrium. His central argument was that unlike the neoclassical model, saving and investment were not necessarily the same thing, in part because different people made decisions regarding saving and investment, and partly because, in a credit money economy, banking policy played an important role; but there was no mechanism to ensure that the quantity of credit available was at a level required to keep saving and investment equal. By the end of the seven years it had taken Keynes to write his *Treatise* however, he had come to believe that while monetary conditions reflected monetary fluctuations, because the fluctuations originated in the realm of business psychology this should be the real focus for investigation (Skidelsky 1992, 314-5).

The General Theory of Employment, Interest and Money, published in February 1936, was Keynes' attempt to reconcile the importance of psychology with the neoclassical framework. Despite the potential slipperiness of observability of psychological factors, Keynes goes to great pains, and largely succeeds, in focusing on relationships between observables. The book itself, though holding an unassailable place at the head of the cannon of twentieth century economics, is as much about introducing a new way of thinking into economics, as it is about setting out a well defined theoretical framework. Indeed, its influence is because of this, and illustrates that although rigorous logic has a key role to play in economics, any logical framework requires application and interpretation. And this in turn requires a less well defined, but nevertheless consistent and believable, way of looking at the world of economic activity.

The richness of the work has led to a plethora of interpretation, but before examining some of these, five features will be set out that are identified by Blaug as the principal distinctive features of Keynes' theoretical framework. These concern the level of analysis of the theory, distinctions between stable

and volatile aggregates, distinctions between saving and investment, interest rate determination and the relationship between wages and employment (Blaug 1997, 646).

Keynes' level of analysis involved a shift from his predecessors in several dimensions. None of the new perspectives was fundamentally new, but each challenged important assumptions of the neoclassical model. It was for this reason that Keynes called his theory general, because he claimed that the neoclassical model made certain assumptions which he did not, and which did not correspond with the available evidence of economic activity, particularly when applied to an economy as a whole (Keynes 1936, xxxii). In this process, as previously mentioned, the primary shift was from what is now termed the micro to the macro perspective. Alongside this was the shift to the short-term equilibrium at the macro level. Although short-term equilibrium was the standard approach at the micro level, the macro level prior to Keynes was analysed in terms of longer duration periods. However, while the micro perspective essentially considered all goods in an exchange as equivalent, Keynes introduced money into the basic exchange mechanism, with distinct characteristics from other goods. The final shift, and in many ways most revolutionary, was the recognition that prices were 'sticky' and variation of quantities could be at least as important as variations in prices.

A key link between investment and consumption was the concept of the multiplier, introduced by Kahn in 1931. This arises from the fact that, when someone is paid to carry out activity as part of an investment, they will spend a proportion of their earnings on goods produced by others, and save the remainder. The multiplier itself, the proportion of the rise in total income compared to the initial investment, depends on the fraction of income saved. Keynes suggested that as income rises, consumption rises, though by less than the full amount (Backhouse 2002, 231). While aggregate consumption and savings are stable functions of income, investment is volatile. So, because of

the multiplier, the volatility of investment means that overall income, and so consumption and savings, will be volatile too (Blaug 1997, 646).

To a large part the predicted instability is a consequence of the observation that the people who make investment decisions are different from the people who make savings decisions. In principle, Keynes saw the level of investment depending on the relationship between the expected return of the investment and the rate of interest. However, expectations are dogged by uncertainty. In theory, expectation should be evaluated over the lifetime of the investment, which can be considerable. In practice, because of the inherent uncertainty involved, the state of confidence at any time will be the critical factor. The implication of this, however, is that apparently minor changes in news can have significant changes in investment, because the current state is generally assumed to be that which will continue indefinitely. Instability is further increased when investment decisions are determined by speculators, because their decision criteria is not even the current situation, but the direction in which the market is moving in the short term. So any real world change is magnified considerably, resulting over time in great instability (Backhouse 2002, 230).

Although the role played by expectations leads to considerable instability, the interest rate does play an important role in investment decision-making. Keynes recognised that money is not limited in its role in financing transactions; the role primarily assigned to it up until that time. It is also held as a store of value. For example, when financial securities are falling in price they have a negative return, so cash is preferable. In general, this leads to the demand for money depending on the rate of interest. This is Keynes' theory of liquidity preference. It makes clear that the rate of interest is not simply a way of connecting savings to investments (Backhouse 2002, 231).

It is the combination of the three elements of the propensity to consume, the marginal efficiency of investment, and liquidity preference that form a theory

of output and employment. An equilibrium point will exist, and changes to the different factors have to be analysed in terms of the resulting change in equilibrium. For example, reducing wage rates will not increase employment, unless doing so raises the level of effective demand, and Keynes, after examining ways this could happen, concluded that this was unlikely. However, raising the money supply or reducing the rate of interest would increase effective demand, and consequently increase output, and so employment (Backhouse 2002, 231).

The description just set out of Keynes' general theory gives a conventional view of the central logic of Keynes argument. Skidelsky characterises the underlying motivation as being to provide a 'logic of choice, not under scarcity, but under uncertainty' (Skidelsky 1992, 539). Yet the logic is only a part of the picture Keynes paints. Skidelsky makes a distinction between the vision, the theory, and the model. The vision of investment in chapter twelve is different from the theory of the investment process in chapter eleven, and the vision of the rate of interest introduced in chapter thirteen and developed in chapter seventeen is different from the formal statement of the theory in chapter fifteen (Skidelsky 1992, 540). Yet despite what could be seen as a rather confused picture, the impact of the book was unprecedented. Why did the book have the impact it did?

5.2.4 Keynes' Impact

In September 1936, a few months after the publication of Keynes' book, three papers were presented at a meeting of the Econometric Society in Oxford. Each presented an algebraic interpretation of Keynes' theory. Of the three, one, John Hicks' paper, drew a diagram, which very quickly became the accepted way of teaching Keynesian theory, and in most cases was the way it was applied in practice (Blaug 1997, 668). Twelve years later, Samuelson's classic textbook, *Economics: An Introductory Analysis*, opened with a section on the determination of national income on Keynesian lines, and went on to

become the most successful economics textbook of all time (Blaug 1997, 643).

Hicks' model, known as the IS-LM model, from the two lines that cross scissor-like inside the positive axis, provided a way of representing both Keynes new theory and the 'classical' theory. The paper, in which he published his model the next year, 1937, was in fact titled 'Mr Keynes and the Classics: A suggested reinterpretation.' The y-axis is labelled 'rate of interest' and the x-axis, 'income.' Both the IS and LM lines represent sets of equilibrium points. The LM curve shows for combinations of income and interest the point at which demand for money, that is liquidity (L), equals an exogenously given supply (M). The IS curve shows combinations of interest and income at which saving (S) matches investment (I). The power of Hicks formulation was its flexibility. Although the LM curve always slopes upwards and the IS downwards, with the LM curve almost vertical and the IS curve almost horizontal the graph represents the classical case, where there is a fixed money supply, and demand for money is a result of its use in transactions, largely determined by income. In contrast, with the IS curve almost vertical and the LM curve almost horizontal the graph represents Keynes' perspective of the 1930s, where demand for money is mainly determined by liquidity preference for holding money as a store of value, and investment and saving follow directly from income. Within these extremes a whole range of relationships can be shown and, at least in principle, can be found empirically, and so provide well-understood options for policymakers to take action, with supposedly predictable results.

While the impact on the economics profession was clearly apparent, it was less so on economic policymaking. Roosevelt's New Deal had got underway with a vengeance immediately following his election victory in 1932, so it makes no sense to claim that it was a practical consequence of Keynes' theory. In fact, on the contrary, the acceptance of Keynes theory, and its implication of an active role for government in the management of the

economy as a whole, could well have benefited from the perceived success of Roosevelt's programme. Skidelsky suggests that following the publication of Keynes' book, and an enthusiastic welcome, particularly at Harvard, the recruitment of young economists from there to Washington did perhaps play a role in reinvigorating a languishing programme (Skidelsky 1992, 580).

With the coming of war, in 1939, there was little time for policy based on Keynes' work to take effect in peacetime. It was ideally placed though to provide the conceptual framework on which the western allies, at least, fought the war on the economic front. Hicks introduced the equation that has subsequently formed the basis of national income accounting: $GNP = C + I + G$ (national income equals consumption plus investment plus government expenditure). Meade and Stone strengthened the framework by introducing a form of presentation in double-entry terms. From 1941 the United States moved away from the Kuznets and Nathan framework, developed earlier in the 1930s, to the Keynesian Meade-Stone one. Subsequently this framework became adopted as an international standard (Backhouse 2002, 244-5).

The availability of figures that people were prepared to trust transformed the economic conduct of war. Keynes himself introduced the concept of the inflationary gap, to calculate how much could be spent without causing inflation, and this was applied to the UK government's budget of 1940. In the US, national income figures confirmed that Roosevelt's 'Victory Program' of vast increases in military expenditure in 1942-3 was achievable; and it was achieved. The approach set feasible and realised targets that saw military procurement rise from 4 to 48 percent of national income, over the course of just four years; a really quite astonishing achievement (Backhouse 2002, 291-2).

It is easy to see why the experience of the New Deal and the war led to a belief that government economic planning along Keynesian lines could effectively tackle inequities and inefficiencies in any economy. In the 1960s

the Kennedy administration in the US was explicit in its use of a Keynesian approach to achieving full employment. The escalation of the Vietnam War, through the 60s, was not planned however, and the resulting monetary expansion turned into rapidly rising inflation. 1973 saw a reduction in oil supplies by the Organisation of Petroleum Exporting Countries, in an attempt to boost their income through a rise in prices. This, and the disruption to supplies as a result of Arab Israeli war in the Middle East, led to an acute shortage of oil. Oil exporters were suddenly receiving more money than they could spend, with the corresponding consequence of reducing demand in oil importing countries. Inflation and unemployment began to rise simultaneously, leaving Keynesian approaches helpless, as rising unemployment suggested increasing spending, but raising inflation suggested reducing spending. The Keynesian consensus had clearly broken down by the end of the 1970s (Backhouse 2002, 295). The only available competitor was monetarism, which saw control of the money supply as the only policy instrument necessary for, and capable of, managing a successful economy. But by the end of the 1980s that had failed too. The 1990s did not see economic theorists providing any clearly visible help to the most serious economic problems of that decade, including the transition from communist central planning in Eastern Europe and Eurasia, and the reverse rather than acceleration of development in Sub-Saharan Africa. While Western Europe and North America continue to thrive economically, few of the people who work in the economy, or the politicians who take the major policy decisions, have studied economic theory, so there is no reason for it to have any part in that success. In the transition and developing countries, where economists are more widely employed in influential positions, economic success is elusive. This is despite economic growth being a major area for contemporary economic enquiry. There is therefore little evidence that the Cambridge combination of Marshallian microeconomics and Keynesian macroeconomics capture in a meaningful way the logic of successful western economies.

5.2.5 Keynesian Economics and Observability

From the above analysis, it is clear that more than anything else, Keynes' theoretical framework was successful for two reasons. Firstly, that it was open to a straightforward algebraic and geometric interpretation, and secondly that it suggested an approach to economic policy that was in line with the contemporary New Deal and wartime experiences of economic policymakers, many of whom were 'called up' to serve for a limited time, before returning to their academic posts.

Keynes himself was probably more conscious of the need for concise frameworks that could be subject to testing, than were his promoters and successors. His concern with measurable properties was to enable his theory to be tested. He did, however, value numerically measurable properties over simply observable ones. While it is arguably better for predicted observations to be measured on a numerical scale, modern molecular biology shows that it is far from a necessity.

The appeal of Keynes' theory to the economics profession clearly arose from the ease with which it could be taught to students, and its apparent success in guiding policy. This illustrates a similar weakness as institutional economics, where although the profession is keen to promote economics as a scientific discipline, there is a 'rush to policy.' Rather than subject theories to empirical test, the profession values the statement of 'obvious' principles with perhaps not so obvious consequences, but ones that nevertheless are promoted as beneficial policy recommendations.

In many ways, Keynes' analysis has stood the test of time as an explanation of many economic phenomena. It is however clearly not the whole of the story. Indeed, while aggregate measures of the economy are useful indicators, by their very nature they are very blunt instruments through which to implement economic policy. As is recognised particularly in the 'new classical macroeconomics,' they say little about the microeconomic level where

interactions actually take place. So, while providing a useful model of the interrelationship between aggregate measures, it has little to say about how micro level activity would lead to different macro outcomes. In other words, its ability to make predictions in particular situations is inherently limited.

The analysis of Keynes' work also suggests however that, while ease of teaching and apparent policy relevance are important for professional acceptance, they are not the whole story. Issues such as the type of phenomena that are worthy of attention, and the specific problems that are susceptible to observation and explanation, are in practice as important as the scientific logic of the work itself in determining its reception and ultimate contribution. So, as we move in the next chapter to assessing the contribution of the commitment-credit theory, it will be worthwhile to take into account work aimed at understanding the nature of scientific progress in general.

Chapter Six

THEORY IN PRACTICE

This final chapter has the task of assessing the theoretical and practical contribution of the commitment-credit model and its tests, set out in the first four chapters, and compared with previous work, as set out in the previous chapter. Chapter one set out the commitment-credit model and argued that governance was an important but overlooked issue that fell between economics and management. The model made specific predictions, which were then tested out in the subsequent three chapters, and confirmed. The previous chapter described two other attempts to incorporate governance into economic analysis, which have received a considerable amount of attention, although neither of them has been particularly successful from a scientific perspective.

It is my contention that most contemporary economists, particularly those who work at the microeconomic level, have lost touch with economic reality in their pursuit of ever more sophisticated mathematical models, with little regard for the important practical problems that good theory could help to address. A growing body of literature such as (Mirowski 1989; Weintraub 2002) provides a detailed intellectual history of this process. The aim of this project, however, has not been to provide an intellectual history, or even a direct critique of contemporary work. Instead, it proposes an alternative, or at least suggests a possible direction along which a better alternative could in time be developed. The limitations of North and Williamson's work, alongside its continuing appeal, suggest that a theoretical framework that combines a ready applicability to practical questions, with a concise and coherent conceptual core is not such an easy thing to do. Nevertheless, some coherent way of assessing the contribution of the commitment-credit model

needs to be chosen. The approach taken here is to use well-established frameworks from the philosophy of science to compare the commitment-credit model and existing alternatives.

6.1 Scientific Method

The initial motivation for this project, set out in the first chapter, was of the need for a theoretical approach that helped address serious practical problems, and in particular, one that took into account increasingly common observations regarding the important role of governance in economic activity. The crucial role of observation in this motivation, and in the relative weight subsequently to be given to it, was said to suggest that this made it a more scientific approach.

There may be the tendency on the part of some to browse through economic journals and conclude that, because the articles are awash with mathematical symbols, it must be a scientific discipline. This would of course be a serious mistake. Since the origins of human history mankind has displayed a need and a facility for abstract reasoning. The ancient origins of astrology and alchemy, alongside their continuing appeal to some, testify to this. Symbols and abstract relationships pervade their study. Metaphysics has a much longer history than physics, but it is not abstraction alone that confers scientific status. Indeed, the difference between science and metaphysics is not in theoretical abstraction but in its tie to reality, and its ability to, not necessarily without resistance, jettison cherished beliefs if the evidence demands it. Maybe in a bid to escape what they see as the woolly reasoning of other social sciences, economists are in danger of ditching not just the social, but the scientific as well.

The claim that the commitment-credit model is more scientific than much of contemporary economics may seem strange. Although the correlation between mathematics and science held good for the nineteenth and first half of the twentieth centuries, the recent success of molecular biology, using

chemistry as its logical foundation, shows that the connection is not a necessity. But if mathematical logic is not what qualifies a discipline as scientific, what is? This is the question to which we now turn.

6.1.1 Underlying Thesis: Science is Observation before Formalisation

Beyond looking for a way to incorporate governance into economic models, the thesis that underlies the specific model presented here is that scientific progress relies on active observation, and that perhaps the most significant failing of neoclassical economics is the readiness of its practitioners to propose practical solutions to imagined economic problems, rather than concentrate on explaining observations, of both success and failure. Economists delight in taking 'self evident' propositions and using these to work out the inevitable consequences of particular policy choices, for either business firms or governments, in terms suggested by the theoretical framework.

On the contrary, society only benefits from economic theory when economists have developed well understood, and tried and tested, logical relationships. It will not be for economists, as scientists, to propose specific solutions, in much the same way that chemists do not prescribe drugs, physicians do, and pharmacists dispense them; because while chemists can predict with increasing certainty how particular molecules will interact, identifying which drugs can tackle the observed symptoms and whether the side effects will outweigh the benefits is a separate profession, where bedside manner is as important as scientific knowledge. Managers and politicians are the professions that should apply economic knowledge, because their training allows them to recognise relevant symptoms and administer treatment in an effective manner. It may be stretching the analogy too far, but maybe lawyers and accountants perform the role of pharmacists in business and political situations, by ensuring that to the greatest extent possible unintended interactions between policies are avoided, as well as avoiding the worst cases of incompetence or negligence in prescribing.

The birth of modern science is often traced to Francis Bacon, and his careful description of natural phenomena. It is the subsequent interaction between logical explanation and careful observation that produced progress in science, and an unprecedented, and still to some awesome, capability of mankind to understand much of its natural environment, and sometimes the ability to control it in some ways. Useful scientific principles such as the relationship between force and acceleration and the role of DNA in the lifecycle of the cell were far from obvious before looking at the world from a certain perspective made them obvious, and enormously useful. There is little role for passive observation. Physical and biological scientists devote huge amounts of time and thought, and increasingly money, to achieve a novel observational perspective.

The classic paradigmatic scientific activity is the experiment. Experiments on animals are increasingly problematic, as are behavioural experiments on people. More widespread social experimentation, and particularly economic experiments on real businesses and nation states are clearly very problematic, to the extent that this can be cited as the major difference between social and natural sciences. But it is in the creation of new observational opportunities that science makes progress. Even new theories do not become important because they are elegant and sophisticated logical constructions, but because they predict otherwise unexpected observations, and sometimes this means that they 'only' explain existing observations, albeit ones that were previously puzzling. In particular, scientists seek to explain both successes and failures, and it is a theory's power to do both that gives it credibility.

Neoclassical economic theory explains neither success nor failure, at least in part because on the whole its practitioners seek not to explain but to prescribe. The commitment-credit model presented here can explain both, and in particular explains the extraordinary phenomena of economic growth, which has released mankind from most of its previous travails and, in providing the resources for scientific progress, has made tremendous strides

in relieving the tedium and grind of hard physical work, and the ravages of many diseases. Maybe the failure of the neoclassical approach is a result of focusing on science as theorizing, rather than as experiment. The relationship between theory and experiment is not straightforward to untangle, but that is the next task.

6.1.2 Popper's Scientific Logic of Falsification

Karl Popper provided what is generally taken to be the definitive analysis of the internal logic of the scientific method. Others have discussed other aspects of scientific activity, and will be considered to some extent subsequently. However, as the first issue at hand is the internal logic of scientific method, Popper's analysis is presented first. In his view, 'A theory is a tool which we test by applying it, and which we judge as to its fitness by the results of its applications' (Popper 1959, 108). Popper's analysis will now be set out in some detail, before considering how this applies to the comparison between the neoclassical and commitment-credit models.

The starting point for scientific enquiry is the theory or model. How it came about is not a scientific question, it is how it stands up to testing that is the scientific issue (Popper 1959, 31). To be able to test a theory it is important to make the distinction between 'universal statements' and 'singular statements' (Popper 1959, 60). Science as an activity is about the search for 'universal statements,' which are general statements, though it proceeds by making predictions regarding 'singular statements.' The two types of statement are distinct and do not overlap. It is never possible to use singular statements to verify universal statements, partly because it is always possible to either dismiss the specific statement as wrong or add an extra condition to the universal statement. In addition, in terms of formal logic it is also impossible to move from a statement of existence to one of universal applicability. However, a statement of universality can be shown to be untrue through the truth of a singular statement that the universal statement excludes as a possibility (Popper 1959, 41). This applies even when the universal statement

is that something exists, because one could hypothesise its opposite, by saying that the opposite exists nowhere, which cannot be verified but can be falsified (Popper 1959, 70).

It is because the only logical relationship between universal statements and basic statements is this one of falsification that this is the focus of the scientific method. It is this that distinguishes science from other speculative activities, and this translates to the central scientific activity of putting theories to the test. While universal statements and the relationship between basic and universal statements are logical statements that can only be disputed on the basis of logical reasoning, basic statements are an intersubjective concept (Popper 1959, 104). What constitutes an effective test of a basic statement can never be incontrovertible, because there is always the opportunity to dismiss a particular test on the basis that it was not doing what it claimed, or that other factors were more important in bringing about the observations in which it resulted. For this reason, repeatability is an important characteristic of a scientific test, although even with this the acceptance of the validity of a particular test is always a collective judgement of the people involved in the discipline.

Popper identifies four possible logical ways that the basic statements that are consistent with the universal statement of a theory can be subjected to testing, and these form the basis on which theories can be compared. They are: logical comparison of basic statements among themselves; determination of logical form to determine whether the theory is empirical scientific or tautological; comparison with basic statements consistent with other theories; and empirical testing of a statement (Popper 1959, 32-33). The first two are clearly internal to the theory itself. The first involves comparing basic statements consistent with the theory to see if they are mutually compatible. Clearly, where a theory suggests both that one test will confirm a theory, and also that the opposite result of the test would also confirm the theory, the theory is not empirically falsifiable, and needs modification. The second test is whether the

theory has any basic statement the result of which could lead to the falsification of the theory; if it does not it is then just a metaphysical theory and not a scientific theory. The second pair of tests allows theories to be compared. Given two self-consistent scientific theories one can compare the basic statements that could lead to the falsification of the theory. It may be that these basic statements are not compatible, or it may be that the basic statements lead to the same tests. Clearly, what is sought is a basic statement which can be tested in such a way that its results will lead to the falsification of either one theory or the other. Finally, of course, a theory can be tested through empirical application. In general these tests will be most productive if they are directly comparing theories, but this may not always be possible.

Because it is testing that is central to the scientific method, it is falsification not formalisation that is the distinguishing feature of scientific theories. Clearly the theories need to be formulated sufficiently definitively and clearly for the introduction of additional assumptions to be recognisable (Popper 1959, 71), otherwise it will be difficult to bring about agreement on what constitutes a testable basic statement. And it also follows that the more opportunities for falsification the better the theory (Popper 1959, 113). So given two theories, one of which implies all the basic statements of the other and others in addition, it is the one with the additional basic statements that should be the focus of attention. While these statements remain unfalsified it is clearly the better theory. But even if these basic statements prove to be false on testing, by making the falsifying results special cases it will still have more falsifiable statements, and so be better from a scientific point of view. Within theories that are broadly equivalent in the extent to which they can be falsified, it is the higher the degree of universality and the higher the level of precision the better (Popper 1959, 122).

From the logic set out above, it follows that scientific progress is made through the postulation of theories that claim to have a higher standard of falsification than existing theories. Even when a particular theory is falsified, it

is always possible to modify it by adding an extra condition that will lead to a basic statement compatible with the outcome of the test that falsified its simpler predecessor. In developing basic statements that will test the theory, it follows that the focus should be on falsification, rather than confirmation, of the theory. It is this dynamic of postulating theories and then testing them until they are falsified, then reformulating the theories, that distinguishes a scientific discipline from scholarly or metaphysical ones.

It therefore follows that to determine the extent to which the commitment-credit model is better, as a scientific theory, than neoclassical theory, and in particular neoclassical partial equilibrium theory, it is necessary to compare them in terms of their falsifiability. So the next task is to establish the falsifiability of the neoclassical theory. After that the two theories can be compared in these terms.

6.1.3 Neoclassical Falsification

Why choose the neoclassical partial equilibrium model, when it is only one among a whole raft of economic models that could be applied in any given situation? Primarily because it is the first model that is taught in conventional economics classes, so even if it is left behind by the time people emerge as qualified economists, it is one of the most important, if not the most important, in the underlying conceptualisation they develop of their subject.

So what is the neoclassical partial equilibrium model? It claims to establish the relationship between seven concepts, namely price, demand quantity, supply quantity, cost, utility of consumption, short and long run, and production scale. It is perhaps one of the distinguishing features of this theory that it can be expressed in both formal mathematical terms and in more intuitive forms requiring less previous education. Any mainstream textbook will provide a very similar development of the relationship between these concepts and demonstrate that, for example, in the short-run a producer should supply the quantity at which the marginal revenue, defined as price at which the next unit

of supply can be sold, is equal to the cost of producing one extra unit of supply in the short run. This is subject to the condition that the price at least equals the short-run average variable cost at that output level. Some recent textbooks are (Fischer, Dornbusch et al. 1988; Mansfield 1994; Hyman 1997; Colander 1998; McConnell and Brue 1999). This is the theory originally developed by Marshall, the context of which was set out in section 5.2.

This is clearly a concise and precise formulation, it is moreover one where the concepts appear to be able to be applied through the measurement of a single quantity. This obviously adds to the pedagogical appeal of the model as students can be given quantitative data in a form that fits easily into this theoretical framework, and asked to carry out analyses within the terms of the theory that lead to precise quantitative answer, which certainly makes marking more straightforward and less subject to the personal judgment of the marker. But this is not what makes a subject a science. As Popper's analysis clearly demonstrates the criterion on which a scientific theory needs to be judged is to what extent is it subject to falsification.

Our suspicions are perhaps aroused when the first classes in economics are compared with those in, for example, physics or biology. In physics, for example, early subjects studied may be Newtonian mechanics or optics. Simple experiments are carried out, and numerical calculations are easily made to establish the relationship between different measurements, for example force and acceleration, or the angle of incidence and the angle of reflection. It may well come over to the students at this stage that theories are constructed by generalising from particular measurements and that this is how science progresses, but nevertheless the intimate relationship between theory and practice is central to the educational process from the very beginning. Biology is in many ways a more interesting example than physics. Unlike physics, modern biology has a central unifying concept, albeit one not as susceptible to quantitative measurement, namely the cell. Again, students use microscopes to examine a variety of animal and plant material and satisfy themselves as to

the veracity of the claimed role of the cell. Any science teacher worth their salt will communicate that it is each student's responsibility to satisfy themselves of the applicability of the theories they are presented with. In no way should they accept what they are told simply on the say so of their teacher, or some obscure reference to 'common sense' or 'everyday experience'.

Economics is taught differently. While there are a huge variety of economic statistics widely available, and one would imagine it is not that difficult to construct, albeit contrived, exercises of stimulated economic activity, or even some data collection from sources readily at hand such as local shops, economics proceeds from asserting the obviousness and importance of the partial equilibrium framework described above. The central concepts of the partial equilibrium framework appear quantifiable, but are they measurable? Because without measurability there can be no falsifiability.

Price is the one easily measured quantity, all the others are problematic, and particularly problematic is cost. It is possible to define the other concepts in a way that is measurable, although further assumptions have to be brought in to operationalise them, so it is not necessarily the core concepts that are being subjected to empirical testing. No methods of directly observing demand and supply quantity schedules are routinely described. Utility of consumption can conceivably be specified as an order of preferences for an extra unit of consumption given an existing consumption pattern, but becomes unmanageably complex beyond this. Both short and long run, and production scale, relate to measurable quantities of time and output, but in quite what way it is impossible to operationalise. Cost is a clearly understandable concept, but it proves fiendishly difficult to measure definitively even in the simplest situations, as the problems grappled with by early accountants described in section 4.1 demonstrate. So, with price as the only observable quantity dealt with in the neoclassical partial-equilibrium model, it is impossible to construct any falsifiable theory; because that would require the existence of at least one other observable quantity for there to be a relationship between them to test.

Given that it is so problematic to falsify the neoclassical partial equilibrium model, it is therefore questionable to what extent it can be considered a scientific theory. This is not to say that other, widely taught, economic theories cannot be considered falsifiable. But given the central role of this model in most economists' education, and the fact that it remains the core of early economics teaching, and no other models have anything like comparable claims made as to the pervasiveness of their applicability, it is difficult to consider economics, as it currently stands, a scientific discipline.

It would seem that it should not be that difficult for a model to be better as a scientific theory than the neoclassical partial equilibrium-model. However, the lack of models that claim this status, and the enduring appeal of the partial-equilibrium model, suggests that finding an economic model, and a microeconomic model in particular, that can be considered in any meaningful way scientific, is a considerable challenge. It is this challenge that the commitment-credit model claims to have met. The next task is to establish the extent to which it can be considered falsifiable.

6.1.4 Commitment-Credit Model Falsifiability

So, what are the key observable conceptualisations in the commitment-credit model, and to what extent are they open to falsification? Falsification is a combination of two elements, namely clarity and observability. Without clarity it will be very difficult to determine whether or not a statement can be said to be consistent with the empirical evidence. This is the weakness of many alternatives to the neoclassical model, including that of North. They deal in several different concepts and relationships, but in any given situation it is difficult to determine how any basic statement would not apply. This clearly makes falsification difficult, which in turn makes the theory of questionable scientific value.

The outstanding characteristic of the neoclassical model is its clarity. The problem set out above, however, concerns the extent to which the concepts it

deals with can be linked with empirical observations. It should be apparent that the commitment-credit model being tested here is clear, although perhaps not to the same extent as the neoclassical model. This however is not sufficient for a theory to be scientifically useful. It is the observability of the concepts that the theory deals with that is where this theory can be considered superior to the neoclassical model.

The remainder of this subsection sets out how each concept defined in the commitment-credit model can be related to empirical observations. The key concepts are innovation, consumption available to be forgone as investment, a credible contracting environment, a financial market for allocating forgone consumption to innovators able to make credible contracts, output growth, the growth of single economic units, growth in the variety of investment vehicles and, within overall output growth, a cycle in the rate. Each of these will be considered in turn.

Innovation, or at least innovative capacity, is taken to be a given, and is qualitatively observable in the pervasive human characteristic of producing tools to make productive activity easier. It is also possible to conceive of measures of the rate of innovation, although this was not pursued in this study. In many ways this is the psychological principle underlying the economic model here, and to the extent that human cognitive capacity is measured through determining the rate at which abstract problems can be solved it is clearly observable. It follows from this that it is the limitation on rationality and information availability that is the foundation of this theory, while neoclassical theory starts with the opposite assumptions and treats situations where these are violated as special cases.

Maximum potential saving is the difference between maximum consumption possible and the minimum in material necessities to sustain life and current level of production. From this definition it is clear that saving potential exists

in most societies most of the time. In monetary economies this can be measured numerically.

A credible contracting environment is something that is difficult to quantify, but it is in many ways the central argument of this thesis that it is observable, and acts as a predictor of economic performance. While the role of innovation and saving in economic activity is widely accepted, the central role given in the commitment-credit model to a credible contracting environment is not. Much as the role of the cell in biological organisms is not at first clear, and needs the appropriate apparatus and expectation to become observable, so it is with the credible contracting environment. Further work may be able to identify specific consequences of particular elements of credible contracting environments.

Financial markets are both observable and quantifiable. In particular they introduce into the model two concepts widely recognised to be important to economic activity, and central to macroeconomics, but absent from microeconomic models, namely money and interest rates. While money can have a role in simply facilitating exchange, its great asset is its ability to be used as a claim on resources invested in future production, to an extent that more than compensates for the forgone conception plus additional consumption opportunities paid as interest to the saver who forwent consumption. In this way it can be conceived of as the original financial derivative. In a monetary economy with extensive division of labour, both the quantity of saving and rates of interest can be measured numerically.

Output growth is both observable and quantifiable. In a money economy output is exchanged for money, so measuring monetary income is essentially the same as measuring output, and the growth in this, corrected for the general rise in the price level through inflation, is relatively straightforward to calculate. One slight caveat is that a competitive environment means that improvements in the quality or performance of the output may not always be

reflected in prices, and so in increases in income. Historically this does not seem in general to have been a significant effect, and to some extent it can be corrected for.

The growth of single economic units is both observable and quantifiable. At this stage this is mainly in the form of the characteristic properties of the industrial revolution and the managerial revolution respectively, namely the factory system and the large corporation. However, by bringing this observation to attention, further work may enable more systematic analysis and further refinement to the theory.

The growth in the variety of investment vehicles is observable, but more problematic to quantify. Each individual type of contract that varies in any way can be considered distinct, but that is not so much the sense of this concept, rather it is the sharing of risk and return between different parties and in different ways. Contemporary empirical financial research on this trade-off may enable this aspect of the theory to be refined relatively straightforwardly.

Cycles are both observable and quantifiable. They essentially follow from shorter-term variations in output growth within the longer-term trend. Distinguishing between short term and long term absolutely is as problematic here as in the neoclassical model. However, because this is an observable consequence of the model, rather than a condition predicting distinct outcomes, in practice it is not nearly so problematic. The business cycle is a well-established topic of investigation, so may make further elaboration only a case of analysing available data, although for the product cycle this is less the case.

This brief summary has shown that all of the central concepts dealt with in the theory can be straightforwardly related to relatively easily observable empirical facts. So while the neoclassical model only has one such concept, namely price, and cannot have any easily testable statements constructed

about it, this does not apply to the commitment-credit model. Because all the primary concepts are straightforwardly related to specific observables, any logical relationships that the theory implies can be subject to test. Therefore the commitment-credit theory is falsifiable, and so a better scientific theory than the neoclassical one.

6.1.5 Distinguishing Historicism and Historical Data

There are also two potential problems with using the historical approach taken here, though neither of them turns out to be serious. It is however worthwhile to consider them briefly, before moving on to the more general question of the practical distinctiveness of the scientific method, and the consequences of this for the contribution of the commitment-credit theory. The first potential problem is the danger that Popper primarily addresses in *The Poverty of Historicism*. It is attempting to use science to predict the future of society. The second potential problem, which Popper also considers, is using empirical data from one epoch to explain experience in another. Both will now be considered in more detail.

Popper defines historicism as the social science that assumes that historical prediction is its principal aim, through discovering the laws and trends that underlie historical evolution (Popper 1961, 3). Alongside this goes the belief that immediate social relations do not necessarily take the same form in different historical epochs, to such an extent that it is impossible to predict them. What historicism does however suggest is that there are universally valid laws that link together historical epochs. Popper suggests however, that this almost certainly overrates differences between epochs. Moreover, most problematic with this as a proposition is that the distinction between epochs is not something that can be open to testing in specific cases (Popper 1961, 92).

Unfortunately, another problem that has followed from historicist work is the confusion that has developed between historicism and holism. Because

historicism assumes that social systems can only be studied as complete wholes, it has taken some comfort from developments in holism and areas of study such as Gestalt psychology, which investigates how people perceive whole objects, rather than their individual elements. Some historicists even refer to alternative approaches as atomistic, on the basis that they treat individuals with no reference to the society that they inhabit. Yet this is a perverse criticism, and can only be rhetorical, because even atomic physics is actually primarily concerned with the relationships between individual particles (Popper 1961, 76).

Popper further criticises the confusion within historicism between laws and trends. Laws are abstract generalisations independent of any specific situation, while trends are observed empirical relationships. Trends refute laws but, as between any abstract theory and empirical result, this is the only possible connection (Popper 1961, 106). One particularly problematic consequence of the historicist position is that it is impossible to predict the effects of small changes, and it is only possible in practice to influence society by accelerating or slowing down large-scale social trends. This has particularly serious consequences, because large-scale change is always problematic. Not the least reason for this is that, when big mistakes occur, they are very difficult to learn from, in part because the impact of different elements is difficult to assess. The most serious problem in practice however is that, because of the many people involved, any attempt at dispassionate assessment will almost inevitably be impossible because particular conclusions could have significant effects on the whole careers of particular people (Popper 1961, 81).

However, while historicism is deeply problematic, Popper does see an important role for historical evidence in social science. Because, while he disputes scientific justifications for any call for large-scale changes such as those advocated by historicism, and historicists dispute the possibility of any enduring generalisation regarding small-scale social phenomena, for Popper it is because of the continuing need for small-scale reforms that social science

has value. Because, while it is never possible to predict with absolute certainty what the outcome of any institutional reform will be, he does not doubt that there are general principles that impose restrictions on what is possible, and these are open to discovery. And this process of discovery benefits enormously from historical experience. So he characterises his position as being anti-historicist, but not anti-historical (Popper 1961, 41). He points out that this corresponds to the situation in the physical sciences, where despite extremely precise theories, there is still a very limited degree of predictability in any concrete situation in the physical world. Although experiments can be contrived, these are very artificial situations (Popper 1961, 129). One consequence of this, not explicitly pointed out by Popper, is that very particular historical situations will need to be selected when seeking to test specific social scientific theories, but this is no different to the very particular situations created in the laboratory.

Unlike the laws of historical evolution sought by historicists, scientific theories logically apply universally, in time and in space, because they are of an abstract nature; it is in their nature that they are never said to apply only to one particular time period. Although even physical laws could always be stated with the condition that they only apply to the present cosmological period, Popper believes that doing this would demonstrate a misunderstanding of the logical nature of scientific theories (Popper 1961, 94). This does not deny that there may be reasons why theories need to be modified to take account of changes between historical epochs, but the scientific approach is to test the application of theories to particular epochs (Popper 1961, 87). Indeed, the differences in social conventions that exist in different geographical regions make it clear that there is a limitation in testing any social theory in just one region or time period that must be recognised, and this does not make the theory less true, it simply means that in different conditions one must expect at least the possibility of different consequences (Popper 1961, 91).

From this analysis, Popper draws implications for history and economics. While historicists present an interpretation of history, such as 'all history is the history of class struggle' they make too strong a claim for their theories. In contrast, what he terms 'classical historians,' have often erred in the other direction and, although they have necessarily adopted points of view, this may not always have been realised. The solution Popper advocates is that points of view should be explicitly adopted, although these points of view may or may not amount to a theory (Popper 1961, 140). Interestingly, given the analysis previously in this chapter, Popper believes that the success of mathematical economics shows that it has 'gone through its Newtonian revolution' (Popper 1961, 54). Perhaps this can be put down to the apparent success of the Keynesian approach to macroeconomics at the time the book was written, though it may also be that in practice things are not as simple, or at least as logically clear, as Popper would like them to be.

The original targets of Popper's argument were Marxist social scientists. It is in many ways ironic that from a contemporary perspective it is in fact neoclassicism, and its political twin neoliberalism, that are now the most popular historicist approaches to social phenomena. For these free marketeers, every specific situation is too difficult and unpredictable to intervene in. To modern historicist ideology, all that can be done is accelerate progress towards state-free all-market nirvana.

From a legal perspective, Atiyah sees the idea of market equilibrium price as practically equivalent to the medieval idea of the just price (Atiyah 1979, 63). It is a worthy concept in practice, but one that can never be legally, or empirically, determined.

In reality, economic phenomena need to be approached scientifically because there is no free-market nirvana. A scientific approach is needed to economics because future consequences of current changes matter, and because

institutional innovation can bring benefits, but there needs to be some way to assess what likely future impacts are, although one can never expect certainty.

In terms of the commitment-credit theory, it is clear that it is not a historicist approach. Although the evidence used to test it is presented as a large-scale historical development, there is no claim that there was anything inevitable in the evolution of social systems that correspond with the commitment-credit model. The model simply suggests that when institutions occur where the attributes identified in the theory are important, there will be particular consequences, and this is what was observed. As a theory it suggests that where the corresponding institutions develop in the future, similar consequences are possible, though they are by no means inevitable. As in any real world situation, there may be additional more powerful factors at work. But does this caveat of 'other factors' occurring in practice not call into question the existence of any meaningful distinction between science, and any other analytical approach? This question will now be addressed.

6.2 Social Science

Although science clearly has an important logical aspect, it is very much a social activity. Subsequent work in the philosophy of science has focused far more on this aspect than Popper did, and two widely regarded approaches will now be considered to help understand science as a distinctive human activity, and the consequences of this for the particular contribution of the commitment-credit theory as science.

While Popper's formulation of the scientific method identifies the essential element that distinguishes science from other areas of human endeavour, it has been recognised for many years that there are other aspects of scientific progress, albeit ones that are common to human endeavour in general. Indeed, Popper recognised the crucial role played by professional convention, specifically in what is accepted as a repeatable test of a prediction. Popper recognised that there was no prospect of a purely logical connection between

abstract theory and the truth of a prediction of a specific relationship between observations; there would always be the need for an element of human judgement.

6.2.1 Kuhn's Metatheory

The most important explanation of the more sociological aspects of scientific progress is that of Thomas Kuhn in his *The Structure of Scientific Revolutions*. His analysis will largely provide the structure for the discussion in the second half of this chapter, which focuses on the potential for a broader contribution of the commitment-credit theory, and the tests of it, set out in this thesis, within the economics profession.

Kuhn's primary concern is with the relationship between abstract theory and professional activity. Although not the only interpretation, this will be taken as an alternative perspective to Popper's definition of science, rather than a direct criticism of it. Indeed, Kuhn implicitly confirms this interpretation because at the end of his first chapter he asks the question 'How could history of science fail to be a source of phenomena to which theories about knowledge may legitimately be asked to apply?' (Kuhn 1996, 9). So he sees himself dealing with abstract formulations that can be tested using historical evidence, much as this work claims to be doing. And because his theory concerns the nature and practice of scientific theory, it seems appropriate to label it a metatheory.

Kuhn therefore, like Popper, has an abstract logical framework for describing scientific activity. He applies this to particular historical evidence, although he is not explicit about testing hypotheses. This metatheory will be set out in some detail, and then applied to the theoretical frameworks described previously. So this is an example not of the testing of a theory, but the application of an accepted theory to make an argument regarding the relationships between the phenomena it deals with.

The perspective taken by Kuhn is of science as a social activity. He defines a scientific community as one that has answers to questions such as 'what are the fundamental entities of which the universe is composed? How do these interact with each other and with the senses? What questions may legitimately be asked about such entities and what techniques employed in seeking solutions?' (Kuhn 1996, 4). Central to the answer the community gives to these questions is what Kuhn terms a 'paradigm.' A paradigm has two characteristics. First if all, it contains a body of accepted theory, which has been applied successfully in such a way that it attracts a group of people to work within the framework to the exclusion of other approaches; but alongside this, it is sufficiently open-ended for there to be unresolved problems for the group to tackle (Kuhn 1996, 10). The concept of paradigm naturally suggests three phases of scientific activity, namely pre-science, before a single paradigm has become established, normal science for work within a paradigm, and revolutionary science, when there is a fundamental shift in the paradigm used by a community. Kuhn's metatheory elaborates these distinctions, which are now set out.

Before a paradigm becomes established, each individual publication starts from first principles, and justifies the introduction of each concept. Research results tend to be communicated using books that are addressed to anyone interested in the subject matter, and non-specialists can follow progress by reading the practitioners' original reports. These phenomena fairly rapidly cease to apply to activity within the field once a single paradigm comes to dominate the discipline, usually coming out of an existing tradition, but one that is initially seen as simply one among a number of possible perspectives on the discipline.

Once a paradigm is established, normal science then occurs. In contrast with pre-normal science, work here takes a good deal as already established, usually presented in educational textbooks, and advances are reported in specialist journals, which are generally difficult for a non-specialist to penetrate. Within

normal science, Kuhn identifies the fact-gathering activity as particularly distinct. This can be seen as largely limited to doing three things. First of all, it aims to increase the accuracy and scope of significant factual observations. Secondly, it seeks to determine facts as yet unknown, but predicted by the paradigm theory. The third aspect involves the collection of facts to aid in the resolution of ambiguities and problems that have been identified within the theoretical paradigm. Kuhn describes the first activity as that to which the greatest amount of effort is expended, and for which great ingenuity is required to push the boundaries of accuracy and application ever further. The second activity, perhaps classically Popperian, is not so widely practiced, often because the formulation of new practicable tests is rarely straightforward. The third type of activity can appear in retrospect to be challenging the theory, but often its initial motivation is just to fill in a gap in knowledge, or establish which of a number of possible perspectives is the best one to apply in a particular situation. Theoretical work is similar, except for the relative proportion of effort that is expended. Some clarification of existing theory and application to new areas is carried out, but is generally regarded as technology not science. Most theoretical work focuses on small and often superficially indistinguishable changes to existing theory, but in a way that is amenable to empirical testing. Theoretical and empirical work most closely relate in the third category, where in practice both advance hand-in-hand (Kuhn 1996, 25-34). One consequence of the paradigm is that, within normal science, fundamental novelties are often suppressed, because they are necessarily perceived as subverting its core commitments. Nevertheless, because those commitments are to some extent arbitrary, the nature of normal research ensures that useful novelties cannot always be suppressed for long (Kuhn 1996, 5). However, in the meantime, a project whose outcome does not fall in the expected range is often seen as a reflection not on nature, but on the scientist conducting the project (Kuhn 1996, 35).

From time to time, it becomes apparent that there are serious problems with the paradigm within which normal science is conducted. Often these arise in

Kuhn's third category, where what can start as a curious anomaly comes increasingly to be regarded as a serious problem (Kuhn 1996, 67). Although this logically indicates a problem with the theory, on the basis that this is the aspect open to human manipulation, theories are not in practice declared invalid until an alternative candidate is available to take the place of the existing one (Kuhn 1996, 77). Kuhn explains this by identifying the interrelation of theory and data as fundamental to the nature of science, and rejecting a theoretical framework without a substitute would mean that potential participants could not regard any resulting activity as science. Furthermore, the very nature of science as a continuing tackling of puzzles that is the hallmark of normal science means that any anomaly can always be dismissed as simply another puzzle that has not yet been resolved, but with time, effort and sufficient brainpower can be solved by using the existing paradigm (Kuhn 1996, 79). Nevertheless, as anomalies continue to persist, competing articulations are put forward; there is an increased willingness to try anything; and explicit discontent begins to be expressed. Debates over fundamentals and recourse to philosophy all provide symptoms of an underlying problem (Kuhn 1996, 91). Then someone very young or new to the field puts a fundamentally new invention forward, and claims to solve the problem that the existing paradigm was unable to tackle (Kuhn 1996, 90, 153). This claim is however rarely sufficient by itself. What can be particularly persuasive is for the new paradigm to predict phenomena that were entirely unsuspected while the old one prevailed (Kuhn 1996, 154). Sometimes aesthetic aspects of the new theory persuade some, though normally only a few, but this group can become the core of an ultimately successful challenge (Kuhn 1996, 156). In the end, the move to the new paradigm is an act of faith, because it promises the potential to tackle many large problems that confront it, even though the older paradigm only fails in a few and has a considerable body of existing uncontroversial material as a result of its previous successes (Kuhn 1996, 158).

At first sight Kuhn's metatheory appears strikingly different to Popper's. It shows signs of embodying a paradigm shift. Possibly for historians and philosophers of science it did, because it substantially broadened the issues it was legitimate for them to consider. Nevertheless, as a framework for assessing the contribution of the commitment-credit model, and its tests, it is in many ways simply a different perspective on the same phenomena. Kuhn does assert that the historical study of scientific development has not found something resembling what he calls the stereotype of falsification by direct comparison with nature, because one paradigm is only rejected when it can be replaced by another (Kuhn 1996, 77). Perhaps the crucial word here is 'stereotype', because this is not a contradiction with the way Popper uses the idea of falsification. The first criteria Popper establishes for a scientific theory is falsifiability. This is essentially taken for granted in Kuhn's framework, where theory development and empirical testing go hand in hand. And as already noted, Popper also makes the point that the actual practice of falsification is a social not a logical criteria, because there are too many factors involved for a statement for or against falsification to be logically definitive; it is a judgement made by a community of scientists, and so different communities, and even the same community at different times, can come to different conclusions. Kuhn's explicit critique of Popper is that Popper makes the choice between paradigms seem too easy (Kuhn 1996, 146). But Popper does not make a distinction between normal science and revolutionary science. Kuhn's description of the vast majority of work in normal science as being refining and stretching standard measurements for experimentalists, and for theoreticians modifying theories in a small way that will be amenable to test through existing techniques, matches with Popper's conception, as does the role of anomalies leading to significant changes in scientific paradigms. In a way, Kuhn's framework is a refinement of Popper's, because while for normal science small changes can be absorbed in the paradigm, sometime changes regarded as necessary by some in the community cannot be so easily incorporated, and a whole-scale rearrangement of conceptual relationships

becomes necessary. So while Popper provides the internal logic, Kuhn provides an external perspective that explains the sociology.

6.2.2 Neoclassical Economic Theory as Kuhnian Science

Kuhn's theory provides a detailed model of scientific progress that has been widely applied to different disciplines, and in some cases further refined. The purpose here is, however, to apply it only to assessing the contribution of the commitment-credit model, and specifically by comparing its potential contribution to the economics profession in relation to the neoclassical, new institutional and Keynesian paradigms.

At first sight neoclassical economic theory displays many of the characteristics that Kuhn ascribes to normal science, but on closer examination it also fails to display others. Economics, and microeconomics in particular, has clearly moved beyond Kuhn's pre-scientific stage, where assumptions are justified each time they are employed, and professional publication is in books rather than scholarly journals. It is however questionable whether what it has moved on to corresponds to Kuhn's concept of normal science.

To some extent a number of threads have to be pulled together to make the problem clear. Kuhn highlights several elements in the education of the neophyte scientist, but two are particularly important. One is the role of textbooks, which Kuhn focuses on largely to demonstrate the fact that previous revolutions tend to be glossed over. Each revolution requires the rewriting of textbooks, or more likely new textbooks, and these new texts present all their material in terms of the new paradigm. Often they present results, theoretical or empirical, from prior to the paradigm shift, but the text does not recognise the different context in which they were originally understood. This reinforces the message that the current paradigm is the 'correct' way of viewing the world, and it is not open to any reasonable challenges (Kuhn 1996, 136-137). The second aspect of scientific training, and the one that to a large extent influenced the choice of the term paradigm, is

the role of exemplary problems and applications that the theory, in order to become accepted, has to better explain than any other candidate (Kuhn 1996, 46-47). Science's pedagogy entangles discussion of the theory with demonstrations and remarks on exemplary applications (Kuhn 1996, 80).

Pulling these strands together prompts the question: what are the exemplary applications that are contained in textbooks that present neoclassical price theory? The answer is, that there are none. A random sample of four microeconomics texts, all from editions published in the 1990s, will serve as a demonstration (Mansfield 1994; Hyman 1997; Colander 1998; McConnell and Brue 1999). McConnell and Brue even describe their text in the preface as the best selling in the United States (1999, xvii). All present in similar ways the price theory framework, use a wide variety of justifications, and no two examples in common, certainly no 'original test' of the theory that, albeit retrospectively, explained the success of the theory. Indeed, there is an emphasis on the revision of examples to bring them up to date (Colander 1998, vii; Hyman 1997, vii). To make their exposition particularly interesting McConnell and Brue even 'present interesting applications with attractive modern art' (1999, xviii). While this combination of art and science is not necessarily a problem, what is problematic is Colander's report of collaborative research he carried out into the opinions of students at 'top graduate programs,' where only three percent agreed that a thorough knowledge of the economy was important to achieving success as an economist, and sixty eight percent said it was unimportant (1998, v). This is particularly troubling when, for example, 'physical intuition' is a phrase commonly used by physicists in explaining the way they work to produce theoretical advances.

Price theory was, as described in subsection 5.2.1, essentially the creation of Marshall, although that attribution is rarely made in textbooks. That is however not necessarily different from textbooks in physical and biological sciences. However, as pointed out, the success of the theory was in

reconciling in a formal way the previously conflicting approaches of Ricardo and Smith. If Marshall did present any exemplary applications, they have disappeared in the numerous generations of successive textbooks. This is in marked contrast to Ricardo's Corn Law application, and Adam Smith's pin factory, which is a part of the broader understanding of economics, although not explicitly taught except in history of economic thought classes.

So although, superficially, neoclassical microeconomics may appear to have the characteristics of Kuhn's normal science, it lacks one key element, namely the paradigmatic application. Of course, as demonstrated in subsection 6.1.3, it lacks any connection with applications because none of its key elements corresponds to any empirically observable phenomena. This makes it virtually impossible for the tension Kuhn effectively takes for granted between theoretical formulations and empirical observations to exist. In particular, the anomalies between observations and paradigmatic predictions that lead to scientific revolutions will be even more infrequent and less apparently problematic, because there simply is not that opportunity for incorporating precise empirical observations into the paradigm. This perhaps explains the enduring role of the framework in economics education. Nevertheless, the inability of the framework to deal with genuine economic problems is increasingly widely recognised, as set out in section 1.1.1.

The next subsection uses Kuhn's framework to answer the questions why the new institutional alternative has not received wider acceptance, and then why Keynes' theory did in the macroeconomic arena, although the subsequent mismatch found between theory and evidence have led to a dramatic loss of popularity.

6.2.3 Neoclassical Alternatives as Kuhnian Science

A cursory review of the new institutional work set out in section 5.1 shows that, because the majority of publications are in the form of books rather than professional journal articles, this alternative to the neoclassical framework is at

a pre-scientific stage of development. It is noticeable and significant that the only work cited that was published as a journal article is by far the oldest, namely Coase's 1937 paper. Presumably at that time his approach was compatible with most other contemporary work, and was seen as mainstream work. However, the other works were all books published in the 70s, 80s and 90s, and by then the fault lines were too pronounced. Each of the approaches will briefly be analysed using the Kuhnian framework, in chronological order.

Coase's paper can be seen as an attempt within the neoclassical framework to develop an explanation for the growing prevalence of large firms apparent by that time, which is not what the neoclassical framework predicted. Of the two dimensions identified by Coase on the basis of which the quality of a theory could be assessed, namely analytical tractability and correspondence with the real world, it is unfortunate that the profession went wholly down the former road, abandoning the latter.

Williamson recognises that his work lies outside the paradigm of normal neoclassical microeconomics, although he does not justify his assumption just in his own terms, but as in line with Coase. Much like Coase, he identifies an anomaly within the neoclassical framework, which asserts the inefficiency of non-market transactions, in contrast with the demonstrable competitive success of many large vertically integrated corporations. His solution is to abandon the central idea of the neoclassical model, namely of economic activity as exchange, in favour of economic activity as contracting. And the competition that leads to efficient exchange is problematic because, while it can deal with bounded rationality and opportunism, it fails to deal with asset specificity, which requires a dynamic treatment outside the instantaneous equilibrium to which the neoclassical framework is limited. Much of the work is however directed to policy formulation, not testing even in the form of confirmation, never mind the possible falsification of the theory. This is perhaps the reason why, although attracting a good deal of attention, an extensive literature has failed to grow up elaborating and further testing the

theory, certainly not to the extent that microeconomic textbooks, although they increasingly include game theoretic frameworks that are not necessarily directly compatible with the neoclassical price theory, also fail to include more recent transaction cost perspectives. Although perhaps this is because, while game theory is not seen as directly contradicting price theory, transaction cost theory is. So Williamson's paradigm does not seem to be sufficiently attractive to replace the existing neoclassical one.

North's work provides a curious contrast between his earlier work, particularly that in conjunction with Thomas, and his later work. The early work provides a close integration of detailed economic evidence and economic theory. The theory is presented as not significantly different to the neoclassical framework, and easily accommodated within it. So it fits neatly into Kuhn's category of normal science, albeit with the discrepancy of appearing as a book. His most recent work however, contends that transaction costs in addition to the neoclassical framework, with its focus on production costs, are not sufficient to explain many economic phenomena. He provides a relatively elaborate articulation of the interaction of motivational benefits and institutional constraints that he claims does provide a significantly improved explanation. Although he does provide a sprinkling of examples, such as national constitutions in North and South America, and general developments in financial markets, there is no concerted in-depth presentation of even one significant economic development, such as that provided in his earlier work. So this perhaps explains why his work, although recognised as important through the award of a Nobel Prize, and while a challenge to the neoclassical framework, has failed to replace it.

Keynes' innovative framework was successful in replacing the neoclassical one, at least in dealing with economic phenomena that have subsequently come to be labelled macroeconomics, and was probably the only large-scale scientific revolution in twentieth century economics. So, before considering the application of Kuhn's framework to the commitment-credit theory, it will

be instructive to examine in more detail the extent to which it applies to the progress of the Keynesian revolution.

Perhaps not surprisingly, given its tenuous relationship with empirical evidence, the anomaly Keynes' theory explained was not focused on one aspect of the neoclassical paradigm, but on its complete inability to either explain or suggest effective policies to deal with the deep and persistent economic problems that affected first of all the British economy in the nineteen twenties, and then the global economy in the nineteen thirties. These problems were exacerbated because the neoclassical framework largely suggested that there were no effective approaches to remedying large-scale economic problems available to a government. And, as Kuhn predicts, it is only when a viable alternative is proposed that moving away from the existing framework becomes conceivable to the vast majority of the profession.

Keynes' status in the profession, as a Professor at a leading university and the editor of the leading professional publication, the *Economic Journal*, is not in line with Kuhn's prediction that challenges come from young or new members of the profession. Although it is perhaps significant that Keynes did not train initially as an economist, but as a mathematician, and then worked as a civil servant, before taking up a university post. He also had considerable subsequent experience in government during the First World War, as well as continuing contact with business through his personal investment activity. What is perhaps significant is that, while as editor he could clearly have published his work initially in the *Economic Journal*, he chose a book as a more appropriate format.

As well as providing an alternative explanation of the contemporary economic crisis, and explaining why Roosevelt's interventionist policy had a significant degree of success, Keynes' theory also provided a significant scientific advance. Keynes went to some trouble to explain that his theory had been formulated in such a way to allow it to be tested using the newly available

national income statistics. Moreover, in doing this he provided additional impetus for advancing this empirical work, as well as providing what was generally regarded as an improved analytical structure for the data. While Hicks' elaboration provided a way of reconciling Keynes' theory with the neoclassical framework, this corresponds to Kuhn's prediction that successful paradigms usually incorporate much previous work, albeit in a different overall framework. So this reconciliation does not diminish the argument for a genuine Keynesian scientific revolution, it actually strengthens it.

6.2.4 Commitment-Credit Theory as Kuhnian Science

Before moving towards a conclusion as to the distinctive scientific contribution of this project, it clearly makes sense to analyse it in some detail in terms of Kuhn's sociological theory of science, as has been done for existing theories in the preceding sections. It will be assessed against each category in turn.

From the perspective of the pre-scientific category, there is some correspondence, but also significant differences. The work is presented in essentially book form, though to some extent that could be simply a function of its role as a PhD thesis. It is also likely that the book could be read by non-specialists and, with the exception of a handful of nuances and references, it is hopefully not too optimistic to believe it could be appreciated. However, the principal concepts are not justified in their own right, but draw on specific, well-established concepts.

In terms of normal science, there are some similarities, but more discrepancies. Aside from its current presentation as a book, it is difficult to envisage a 'core' part being published in mainstream journal, without removing so much material that it would detract from the overall argument. There is little sense in which a professional audience could take as read extensive references to the background literature and detailed methodological concerns, which normally disappear in the transition from PhD thesis to

published paper, and which are largely presented in the former to demonstrate the professional status of the author, in addition to the small but recognisable advance at the core of the work. The primary innovation here is in the theory at the beginning, though it is the weight of significant empirical evidence presented subsequently that justifies its value; and there is no claim to originality in the empirical evidence, though in fact this provides the bulk of the work. But, with the possible exception of North and Thomas, this is not a conventional form of empirical economic work. So it would be difficult to characterise it in Kuhn's term as a theoretical advance that does not move on too much from existing theory, because its application to empirical data, in such a distinctive way as is done here, would be very problematic. The form of the work, although broadly similar to that of Williamson and North, is very different in form from that which currently fills most professional economic journals. So the closest correspondence to a form of normal science is the exploration of an anomaly between the lack of growth mechanism in neoclassical microeconomics, using the very much minority but nonetheless existing approaches of theory expressed abstractly in verbal form, drawing together financial economics and the property rights aspect of the new institutional framework, together with extensive historical evidence. The closest similarity is with the work of North and Thomas of three decades ago, so is hardly contemporary.

So the final stage is to consider the extent to which there is evidence for a correspondence between this work and Kuhn's notion of a scientific revolution. The initial indications are that there is a degree of agreement. Work combining empirical and theoretical innovations to explore anomalies within existing theory and data is the normal source of ideas that are able to develop to the extent that they lead to revolution. But, as occurred with the transaction cost and new institutional economic theories, although they have explained empirical phenomena that the neoclassical framework has difficulty in admitting, particularly the growth of large corporations and their enduring success in competitive markets, this work comes up against the problem that

the neoclassical framework has no close connections with any empirical data, so it is difficult to argue successfully from a discrepancy between theory and evidence the need for an alternative paradigm.

However, the anomaly that the commitment-credit theory primarily addressed is not one of explaining success, which can perhaps always be countered by saying that alternatives could be even more successful, but is a sustained crisis in economic experience on a scale at least comparable to that which preceded the Keynesian revolution. This is the increasingly recognised widespread failure of both neoclassical and Keynesian frameworks in explaining and suggesting enablers of economic growth, for what is often, somewhat euphemistically given the evidence, called the developing world. However, as Kuhn's theory and the evidence of the Keynesian revolution strongly suggest, this need is only acknowledged in hindsight, once a credible alternative paradigm has been developed.

One indication predicted by Kuhn of the coming of a revolution is recourse to philosophical debates. It is not clear from the evidence presented here that this was present to any real extent in economics prior to Keynes. Although it is not necessarily a coincidence that, prior to economics, Keynes' primary intellectual interest was the philosophical nature of probability theory, and although his theory when formalised did not include any probabilistic elements, chance and uncertainty were significant aspects of some elements of his paradigm. There is some evidence of philosophical questions currently being asked, in work such as (Blaug 1992) and (Mirowski 1989), and the recent call by some for a 'post autistic economics' (www.paecon.net), but these appear relatively peripheral to the discipline as a whole. Nevertheless, the problematic nature of a challenge to an existing paradigm, in that it makes no sense to attempt to evaluate any contribution in terms of that paradigm, is why philosophy of science is the framework drawn on here to discuss the contribution of the commitment-credit theory.

6.2.5 Scientific Contribution of the Commitment-Credit Model

Given the analysis just set out, assessing the commitment-credit theory in its relationship with alternative economic theories, in terms of the metatheories of Popper and Kuhn, what can be concluded as to its contribution?

From the Popperian perspective, the evidence presented in chapters two, three and four, although nominally focused on attempting to falsify the theory, does overwhelmingly show that there is a way of matching the abstract constructs of the model with observable phenomena. Because Popper focuses on the logic of scientific activity, he pays no attention to the social aspects, except where this is necessary to fill a gap in the logic, and the key social aspect is agreement on the correspondence between the abstract and observable phenomena. While for physical and biological scientists this activity of matching theory and data is probably the most substantial aspect of their training, it is largely absent from contemporary economic education. And the claim in the first section of this chapter was that this is the primary difference between the neoclassical and commitment-credit model. So, to what extent has this objective of a microeconomic theory that has clear empirical application been achieved?

The essential pivot of the argument for the scientific value of the commitment-credit theory is that historical observations confirm the prediction of the theory: that the result of developing an effective form of commitment device, which allowed savings to be invested through a financial market, did harness the natural human talent for innovation, so that it was directed towards economic growth.

Although the ancient and medieval Mediterranean world saw the early stages of developments that would lead to effective economic commitment devices, they were unable to provide the confidence necessary to override wider social concerns. It was only in England that the legal infrastructure necessary to arbitrate and enforce contract disputes rationally slowly emerged. Although in

many cases taken for granted today, and despite the difficulties in achieving this often being belittled, the achievement was real and distinct. However, as contractual links increased in quantity and sophistication it became necessary for large legal entities to be able to provide evidence of their ability to live up to their commitments beyond legally enforceable contracts. Audited accounts met this need, and achieved their modern role under the auspices of financial market regulators in the United States in the early twentieth century. It is the nature of these commitment devices that they are written documents potentially open to subsequent legal scrutiny, so the character of their historical development remains available to scientific scrutiny.

Even in ancient times commitment to repayment was a necessary condition of lending, but the difficulties in disentangling economic and broader social relationships prevented the development of financial markets that directed funds to the most economically beneficial uses. Largely as a result of the legal environment, effective impersonal financial markets developed from the end of the seventeenth century in England. Initially this was through discounting bills of exchange, with the national debt providing an effectively risk free form of saving that provided the benchmark against which the risk and return on private liabilities could be assessed. Then an increasingly specialised infrastructure of financial markets and intermediaries developed, to the extent that a qualitative change in their nature occurred, and reached its modern shape in the United States by the early years of the twentieth century. The public nature of these markets has also left its mark on the historical record, so making the character and development of credit markets observable.

While the industrial revolution is a widely recognised phenomenon and, although not without bringing significant social problems, is generally seen as providing massive economic benefits, quite what its nature was and the reasons for its success are less widely subject to consensus. And although large organisations are pervasive features of modern life, their origin and indeed sometimes their role in society are less clear. The four aspects of

growth in the commitment-credit model, namely in output, in sub-unit size, in credit sophistication, and in cyclical form, provide a relatively clear framework in which developments can be assessed. Although not necessarily as immediately available to scrutiny, because their observability is not as central to their function as it is to commitment devices and credit markets, they have all left their mark. Output measures have been a subject of intense activity for economic historians over the years, particularly at the national level before the advent of official statistics, and these statistics provide at least a good first estimate where they are available. Financial accounts likewise provide measures for firms. Unit size is perhaps the most problematic, in that systematic data sources do not exist, although details of many particular cases do. Financial sophistication is observable through an examination of the extant instruments, which because they were written contracts are often still available, though again no obvious systematic source or analysis is extant. Cycles again are widely recognised and data is available, though not on a wholly systematic basis.

It is on the basis of the observability of its conceptual elements that the commitment-credit model can claim from a Popperian point of view to be an advance over existing alternatives. Neither neoclassical nor new institutional microeconomics offers this combination of analytical precision and observable falsifiability. Keynesian economic theory does, and although it deals with saving and investment, and its focus on aggregate statistics has been extended to tackle economic growth, it is an unabashed macro perspective and to some extent relies on the veracity of the neoclassical model in dealing with micro level phenomena.

From the Kuhnian perspective the contribution of the commitment-credit theory is more difficult to assess definitively. It does stand relatively sustained comparison with the initial stages of the Keynesian revolution. In some ways it could be seen to carry forward the Keynesian revolution by using a Keynesian approach at the micro level, that is, using what Skidelsky describes

as the vision of the theory in providing a logic of choice, not under scarcity, but under uncertainty. Although while Keynes dealt with concepts made observable through national statistics, the commitment-credit model deals with those made observable through the analysis of historical records. Much as in Keynes' time, where some people in Britain lived in considerable luxury while many others lived in abject poverty, the same is widely regarded as true today on the global scale. And while choices have to be made by governments, by individuals and by firms, the problems to be solved are not generally seen as problems of scarcity, but as problems of uncertainty. The need to do something is evident. The problem is that there is little agreement on what should, or even can, be done.

Many of the concepts and approaches that make up the commitment-credit model also play important roles in the new institutional literature. North and Thomas' book is the most similar part of the literature to the overall approach taken here. And Williamson gives contracts an equally pivotal role in his model, though they are seen as important in production and not in innovation and investment, which are the focus here. His book also commences with a relatively concise abstract formulation, although he then goes on to apply this to current policy debates, rather than assess its explanatory capability when applied to phenomena about which there is a good deal of consensus. So in Kuhnian terms this work could be a contribution to the largely still pre-scientific new institutional economics.

It is the nature of the sociological basis of Kuhn's framework that it is impossible, for now, to assess the ultimate contribution of this work in his terms. It will not be clear for some time whether this work will be an integral part of a revolution in economic science or simply part of the froth outside the mainstream of the profession. Perhaps the most that can be said definitively is that it shows that an alternative is possible.

However, regardless of a theory's philosophical status and its ability to explain an existing anomaly, Kuhn is clear that an alternative in itself is not sufficient, because it must explain phenomena not previously explained. Although this may be a useful criterion in the natural sciences, it is not so useful in the social sciences, where particular events are always explainable to some extent in terms of their social and historical environment. Perhaps it should be that the model naturally explains phenomena in several situations that are otherwise only explainable in terms of their specific situation. Nevertheless, in the end the acceptance of a new paradigm is an act of faith by a substantial community of scholars, faith that the new approach will be able to deal with problems existing theory cannot, and only time can hope to provide the answer as to whether the commitment-credit model will achieve that.

So the commitment-credit theory provides an alternative, but as with the Keynesian revolution, scientific criteria will almost certainly not be enough to attract further attention and elaboration within the economics profession, it will also need to have important practical implications. So, finally, the question needs to be answered as to how this theory differs from a practical perspective. One might even say: it is all very well in theory, but what about in practice?

So, the commitment-credit model has some claim to have made a contribution to economics as a scientific discipline, but does it have any practical implications? While it could be argued from a scientific point of view that conciseness is a big benefit to a scientific theory, are not societies and economies so complicated that in practice no simple model can be of any practical value? Can the scientific method broadly construed ever really be expected to make any useful contribution to social phenomena; is there anything to distinguish its products from those produced from an artistic or contemplative perspective? Applying these questions to the commitment-credit model: how does it contribute to ongoing debates of the relationship between science and practice, and particularly social science and social

practice? This is the question that the final subsection seeks to answer. Since Popper's and Kuhn's writing there have been sustained critiques of the existence of science and scientific knowledge as a distinct activity, although perhaps the concern was that science has in some situations become viewed as a privileged form of knowledge, and perhaps this is problematic. This critique in effect questions whether there is any meaningful way that theories, including the commitment-credit model, can be practically compared. One popular critique of the concept of a distinct scientific method serves as the focus here, that of Feyerabend.

6.2.6 Critique of Scientific Distinctiveness

Feyerabend in his book *Against Method* under 'the principle: anything goes' (Feyerabend 1993, 19) argues that it is impossible to set out any rational principles that must always apply to any scientific activity. He gives the example of Galileo, who can be seen in retrospect as having a deep and profound effect on subsequent science, but who worked in what seems to be a highly unsystematic way. Galileo's results were also problematic. For example, while use of the telescope he invented on terrestrial objects was acknowledged as unproblematic, he had considerable difficulty with celestial phenomena. For example, at one demonstration to twenty-four professors they could not agree that they were able to make any useful astronomical observations (Feyerabend 1993, 88). The experimental comparison this was supposed to give, by current standards, would not have been able to establish a distinction between the Aristotelian and Copernican alternative theories. Modern theories of, for example, the Coriolis forces that influence of the earth's atmosphere, optical aberrations, after images, and mutual inhibition of retinal elements, now explain this difficulty; but were obviously far away in those days (Feyerabend 1993, 110).

The implication of the historical difficulties Galileo experienced is, for Feyerabend, that Popper's theory of science does not predict these difficulties. Popper's theory 'is either a meaningful idea or it is a collection of slogans that

can be adapted to any situation' (Feyerabend 1993, 151). The irony is that this applies to Feyerabend's 'argument', although it is difficult to tell to what extent this is deliberate. In the book's preface, where its overall approach is set out, Feyerabend says that he uses irony and mentions 'anything goes' in particular as not so much a principle but 'the terrified exclamation of a rationalist who takes a look at history' (1993, vii). Later in the book Feyerabend describes the difference between his and Kuhn's work in that his is abstract while Kuhn's is historical (Feyerabend 1993, 212). A better description would seem to be that Kuhn's is historical and abstract, while his is just abstract. Even better, both are abstract and historical, but Kuhn proposes an abstract characterisation, while Feyerabend does not.

While Feyerabend seems to be arguing against science, he is in fact arguing against philosophy; what Popper would call metaphysics. It does indeed seem impossible to find general statements that apply in all situations, unless they are so woolly and general that they are effectively devoid of meaning. But this does also beg the question of the value of Popper and Kuhn's theories, and the idea of science as a distinct activity. But some considerations will make clear that, unlike Feyerabend, these do have value beyond that of a brain-teasing puzzle.

Feyerabend's basic problem is one that seems to be more at home in the nineteenth century, though perhaps he would argue that, on the contrary he is arguing against a nineteenth century view that was still enduring in the twentieth century. Even before the end of the nineteenth century, in 1895, Poincaré had shown that there was not a fully deterministic solution to the 'three body problem' of the gravitational interaction between three otherwise independent bodies. So a precisely specified abstract system has an indeterminate outcome. The quantum revolution in atomic physics, and Gödel's proof of the impossibility of a complete logical system without contradiction, in 1931, simply reinforced the idea that from both a physical and logical point of view it was impossible to reconcile science and

determinism. Why after this anyone would wish to argue for any broadly applied and absolutely correct generalisations is difficult to comprehend. The complexity and indeterminate nature of the human interactions at the core of any social process is no excuse for denying the possibility of taking a scientific approach to understanding them. The issue is perhaps to get beyond the simplistic conception of science and scientific theories.

Scientific endeavour itself, it almost goes without saying, is a human social activity. Popper recognises this, as has been repeatedly referred to previously, through his recognition of the consensual nature of recognising particular observations as confirming or falsifying basic statements. Feyerabend's description of Galileo's initial difficulty concerning celestial observation with his telescope clearly falls into this area. And Kuhn's whole approach takes the social nature of science as its starting point. While Kuhn introduced the idea of paradigm shift, he clearly did not introduce the idea of revolution. Indeed, in many ways, Kuhn's characterisation of scientific revolutions as paradigm shifts sheds a great deal of light on other social changes, including the legal, financial, industrial, accounting, corporate finance and managerial revolutions described here in chapters three and four.

Feyerabend's point is perhaps that science is no different from any other human activity. Perhaps more strongly, that it is not a privileged form of human endeavour. It is indeed problematic to describe it as privileged, but that is not a claim made here. It does seem particularly problematic however to say that because no simple generalisation of the scientific method can be articulated then science is not a distinct activity. Kuhn, and perhaps more so Popper, articulate theories that do seem to capture the key elements of scientific activity, in that they are concerned with the interaction between formal knowledge and empirical data, largely for their own sake, and this is what distinguishes scientific activity. But as Kuhn recognises, a paradigm such as his does as much to create problems as it does to solve them. Popper captures the idea that more general theories create more challenges from a

wider range of specialisms, and this is in line with what intuitively makes a better theory.

So it does need to be said that it is inherently problematic to try to discuss practical consequences of scientific development. Science can only be effective when it proceeds on its own terms. Nevertheless, it is an integral part of the wider world, and it is even more problematic to simply treat it as an autonomous activity. So the final question answered here is, what could be the distinctive practical contribution of commitment-credit theory as presented here?

6.2.7 Practical Contribution of Commitment-Credit Theory

For most of the twentieth century there was an increasingly loud call for the separation of economic theory and economic policy, articulated as the need to distinguish between positive and normative economics, with the more or less explicit corollary that only positive economics can claim to be scientific. It is also a striking coincidence that the best-known proponents of this distinction, Lionel Robbins and Milton Friedman, were also amongst the most politically conservative economists of their generation.

Lionel Robbins in his 1933 book *The Nature and Significance of Economic Theory* argued that economics was the scientific study of choice. It was here that one of the most popular and enduring definitions of economics was set out, as making choices under conditions of scarcity. Although he argued that this was the positive, and therefore the scientific approach to economics, he argued that it could be pursued largely independently of empirical work (Backhouse 2002, 239). Blaug cannot bring himself to accept that Robbins really believed that theories should be developed simply through deducing the logical consequences of simple postulates derived from introspection, although this would not be incompatible with anything in the book (Blaug 1997, 690). Nevertheless, this has become how microeconomics in particular is largely

conducted, despite the subsequent rise to prominence of Popper's and Kuhn's work.

Friedman published his 'Essay on Positive Economics' in 1953. It can be concluded from this that he is arguing that predictive accuracy should be the sole criteria of a theory's validity (Blaug 1997, 694). Given Popper's insight that the step from theory to practical observation is the one step not susceptible to logical analysis, and can only be dealt with through a professional consensus, Friedman's proposal is problematic. In the essay itself Friedman does to some extent recognise this, although his argument is more of a negative one against criticism of the 'unrealistic' assumptions of neoclassical economics. Unfortunately, he does not offer any prediction of neoclassical economics that had been confirmed, and simply asserts his belief that it is fruitful and deserves confidence (Friedman 1953, 207).

In many ways positive economics is in political terms a deeply conservative concept, because it posits that what exists is in some sense superior to alternative preferences, and leads to problems such as the importance attached to the Pareto equilibrium. Although this concept recognises that not all policy interventions are zero sum, because some changes could be made that would improve the economic position of one individual while not affecting that of others, it begs the question what is so crucial about no one losing out? Although it can be precisely stated, this is not the only criterion on which theoretical concepts are chosen. Why should it be considered of value within an economic theory? It was precisely the failure to tackle particular vested interests that lay at the heart of Argentina's economic failure. This is not to say that economists should make decisions based simply on their theories, but that their role is to set out likely consequences, and not to reject policies, much less whole theories, that may breach normative principles such as Pareto equilibrium.

Perhaps most problematic with the positivist approach to economics is the 'rush to policy' problem. Putting aside Friedman's exhortation to only accept theories with demonstrated predictive power, Robbins' implication that economic consequences can be deduced simply through a combination of introspection and deduction provides ample justification for rushing to implement policies justified by these untested theories, though normally the real justification is political.

Enron and, through the IMF, Argentina, were guided by a political ideology and not by scientific principles. This is not the place for a detailed consideration of the experience of Enron and Argentina as a test of the commitment-credit model, although this may be an interesting subsequent research project once more data is available, but it is striking that the theory puts at the centre of attention issues such as managerial accountability and legal predictability that are now generally regarded as the reasons behind such spectacular, and to some extent unexpected, failures, yet form little part of contemporary economic research.

Ideology and politics are not in themselves problematic; in many cases they are the only way to gain sufficient consensus to deal with controversial social problems, or to take advantage of poorly understood new economic opportunities. It is when science and politics are confused that problems arise. Although it should be clear that science and politics are different things, this in no way means that they have nothing to offer each other. Feyerabend can be seen to be making a positive contribution by pointing out the absurdity of science 'trumping' politics because it has some innate truth that politics lacks. Politics can draw on scientific work to identify possible solutions to the difficult problems with which it deals, and even to help in arguing the case for a solution that has been chosen. Science on the other hand, and social science in particular, by being engaged with politics, can be guided as to where problems exist that it could tackle with a possibility of contributing to society,

from which ultimately it needs resources. Politics here includes both 'big P' national politics, and 'small p' managerial politics.

The commitment-credit theory is clearly still in its very early stages, and does not offer a great deal of precision; nevertheless, it does offer some benefits to society. The germ theory of disease reinforced a growing focus on the importance of hygiene, long before anything was known about the specific molecular processes underlying particular diseases, which are only now starting to be understood. But Kuhn recognises that a change in paradigm often appears to lose detail. This is clearly the case in microeconomics, where contemporary economic theory sets high standards of precision, and generally requires the use of formal mathematical language. This is a feature, which Popper at least recognised as having value, which the framework as set out clearly fails to meet. This begs the question, to what extent can this formulation be formalised? The arbitrage-free capital market logic is heavily formalised, so this may provide a platform to build on. Although it is not immediately clear to what extent this could be extended to the innovative and productive side of the economy, apart from the capital markets. An important part of the model is the fundamental uncertainty of innovation that the capital markets exist to share out, and so minimise the extent to which individual investors are exposed. However, it may turn out that rather than the likelihood of a particular product's success, statistically it is the number of innovations brought to market that is significant overall, together with the propensity of consumers to try alternative products and services. Consumer choice would act as a ratchet, only repeating purchases of goods and services that provide a higher benefit/price ratio, leading to growth in the economy. Alternatively, it may turn out that a jurisdiction-independent legal-regulatory logic needs to be more carefully articulated, and quantitative measurements will always be more subsidiary. Clearly, further research could produce important advances here.

It could be said that the consequences of the problems with Enron were far less severe because it existed in an effectively regulated environment, but it is also the case that moving to another employer is routine, while moving to another country is not, at least given contemporary restrictions on cross-border labour mobility. So while commitment-credit theory does provide a paradigm that goes a long way towards understanding many of the issues regarding the management of firms, it is not sufficient for explaining the additional complications of national economies. It cannot tell managers what to do, but it does seek to address what they perceive as problems and provide a conceptual framework that helps them to think through the implications of alternative decisions.

What the commitment-credit theory does imply is a shift in economic theorising, one largely inverting micro and macro concerns. Micro's current focus is largely on equilibrium, while macro's is growth. Yet growth is now generally understood to be a result of micro activity, producing not just more but better goods and services, and should be considered at this level. It follows that the role of the macro-economy, and macroeconomic policy, should be to ensure that the micro-level transitions, which innovation implies, are enabled to happen in a way that involves the least disruption across the economy as a whole, and for the individual people who provide its most important raw material. This suggests that further work is needed to build on this microeconomic theory, to create a macroeconomic theory that seeks to understand the additional role for democratic accountability in regulating legal, accounting and financial systems, and in managing the money supply and public finances.

Of course, proposing what is effectively a new paradigm for understanding microeconomics, one that makes far more sense from a managerial point of view than the existing neoclassical one, creates far more questions than answers. That is the nature of a scientific paradigm. Perhaps there is some comfort that one question is better understood. Einstein is reported to have

replied to the question as to what is the most powerful force in the universe, with the answer 'compound interest.' The commitment-credit theory goes a long way to explaining why this makes so much sense, because it puts investment rather than production and consumption at the heart of economic analysis. Only time will tell whether the economics profession can abandon simple exchange as its central paradigm, and in its place institutionalise commitment and credit.

BIBLIOGRAPHY

- Andreu, J. (1999). Banking and Business in the Roman World. Cambridge, Cambridge University Press.
- Atiyah (1979). The Rise and Fall of Freedom of Contract. Oxford, Clarendon Press.
- Backhouse, R. (2002). The Penguin History of Economics. London, Penguin.
- Baines, E. (1835). History of the Cotton Manufacture in Great Britain. Human Documents of the Industrial Revolution in Britain. E. R. Pike. London, George Allen & Unwin.
- Baker, J. H. (1979). "The Law Merchant and the Common Law Before 1700." Cambridge Law Journal 38(2): 295-322.
- Baker, J. H. (1990). An Introduction to English Legal History. London, Butterworths.
- Baskin, J. B. and P. Miranti (1997). A History of Corporate Finance. Cambridge, Cambridge University Press.
- Bernstein (1993). Capital Ideas: The Improbable Origins of Modern Wall Street, Free Press.
- Blaug, M. (1992). The Methodology of Economics: Or How Economists Explain. Cambridge, Cambridge University Press.
- Blaug, M. (1997). Economic Theory in Retrospect. Cambridge, Cambridge University Press.
- Bodenhorn, H. (2000). A History of Banking in Antebellum America. Cambridge, Cambridge University Press.
- Buchinsky, M. and B. Polak (1993). "The Emergence of a National Capital Market in England, 1710-1880." Journal of Economic History 53(1): 1-24.
- Caenegem, R. G. V. (1992). An Historical Introduction to Private Law. Cambridge, Cambridge University Press.

- Carosso, V. (1970). Investment Banking in America: A History. Cambridge MA, Harvard University Press.
- Carruthers, B. (1996). City of Capital: Politics and Markets in the English Financial Revolution. Princeton N.J., Princeton University Press.
- Chandler, A. D. (1962). Strategy and Structure : Chapters in the History of the Industrial Enterprise. Cambridge, Mass ; London, M.I.T. Press.
- Chandler, A. D. (1977). The Visible Hand : The Managerial Revolution in American Business. Cambridge, Mass. ; London, Harvard University Press.
- Cibils, A. B., M. Weisbrot, et al. (2002). Argentina Since Default: The IMF and the Depression. Washington D.C., Center for Economic and Policy Research.
- Cipolla, C. (1994). Before the Industrial Revolution: European Society and Economy 1000-1700. New York N.Y., Norton.
- Clark, G. (1993). Agriculture and the Industrial Revolution. The British Industrial Revolution: An Economic Perspective. J. Mokyr. Boulder CO, Westview.
- Coase, R. (1937). "The Nature of the Firm." Economica n.s.(4).
- Cohen, E. (1992). Athenian Economy and Society: A Banking Perspective. Princeton NJ, Princeton University Press.
- Colander, D. (1998). Microeconomics. Boston MA, Irwin McGraw-Hill.
- Daunton, M. J. (1995). Progress and Poverty : An Economic and Social History of Britain, 1700-1850. Oxford, England, Oxford University Press.
- Davenport (1989). James Watt and the Patent System, British Library.
- Davies, G. (1994). A History of Money : From Ancient Times to the Present Day. Cardiff, University of Wales Press.
- Dickinson, H. W. (1936). Matthew Boulton. Warwickshire, TEE Publishing.

- Dickson, P. (1967). The Financial Revolution in England: A Study in the Development of Public Credit 1688-1756. London, Macmillan.
- Feyerabend, P. (1993). Against Method. London, Verso.
- Finley, M. I. (1985). The Ancient Economy. London, Hogarth Press.
- Fischer, S., R. Dornbusch, et al. (1988). Introduction to Microeconomics. New York, NY, McGraw-Hill.
- Friedman, M. (1953). The Methodology of Positive Economics. The Philosophy of Economics: An Anthology. D. Hausman. Cambridge, Cambridge University Press.
- Fusaro, P. and R. Miller (2002). What Went Wrong at Enron. Hoboken, NJ, John Wiley and Sons.
- Garner, S. P. (1954). Evolution of Cost Accounting to 1925. Alabama, University of Alabama.
- Gleeson (1998). The Arcanum: The Extraordinary True Story of the Invention of European Porcelain, Bantam.
- Harley, C. K. (1993). Reassessing the Industrial Revolution: A Macro View. The British Industrial Revolution: An Economic Perspective. J. Mokyr. Boulder Co, Westview.
- Harris, R. (2000). Industrializing English Law: Entrepreneurship and Business Organization, 1720-1844. Cambridge, Cambridge University Press.
- Hart, O. D. (1995). Firms, Contracts, and Financial Structure. New York, Clarendon Press.
- Hudson, P. (1992). The Industrial Revolution. London, Arnold.
- Hunt, E. and J. Murray (1999). A History of Business in Medieval Europe, 1200-1550. Cambridge, Cambridge University Press.
- Hyman, D. (1997). Microeconomics. Chicago, Il, Irwin.
- James, J. (1978). Money and Capital Markets in Postbellum America. Princeton NJ, Princeton University Press.

- Johnson, H. T. and R. S. Kaplan (1987). Relevance Lost : The Rise and Fall of Management Accounting. Boston, Mass., Harvard Business School Press.
- Johnston, D. (1999). Roman Law in Context. Cambridge, Cambridge University Press.
- Kapr (1996). Johann Gutenberg: The Man and his Invention, Sclar Press.
- Keynes, J. M. (1936). The General Theory of Employment, Interest and Money. London, Macmillan.
- Kindleberger, C. (1993). A Financial History of Western Europe. Oxford, Oxford University Press.
- Knack, S. (1999). Social Capital Growth and Poverty: A Survey of Cross-Country Evidence. Washington D.C., World Bank.
- Kostal, R. W. (1994). Law and English Railway Capitalism 1825-1875. Oxford, Oxford University Press.
- Krueger, A. (2002). A New Approach To Sovereign Debt Restructuring. Washington D.C., International Monetary Fund.
- Kuhn, T. (1996). The Structure of Scientific Revolutions. Chicago, IL, University of Chicago Press.
- Lamoreaux, N. (1985). The Great Merger Movement in American Business, 1895-1904. Cambridge, Cambridge University Press.
- Landes, D. (1993). The Fable of the Dead Horse; or, The Industrial Revolution Revisited. The British Industrial Revolution: An Economic Perspective. J. Mokyr. Boulder CO, Westview.
- Levenstein, M. (1998). Accounting for Growth: Information Systems and the Creation of the Large Corporation. Stanford CA, Stanford University Press.
- Lewis, C. (2002). Argentina: A Short History. Oxford, Oneworld.
- Lopez, R. (1976). The Commercial Revolution of the Middle Ages, 950-1350. Cambridge, Cambridge University Press.

- Mansfield, E. (1994). Microeconomics. New York, NY, Norton.
- McConnell, C. and S. Brue (1999). Microeconomics: Principles, Problems and Policies. Boston, MA, Irwin McGraw-Hill.
- Mirowski, P. (1985). The Birth of the Business Cycle, Garland.
- Mirowski, P. (1989). More Heat Than Light: Economics as Social Physics, Physics as Nature's Economics. Cambridge, Cambridge University Press.
- Morgan, K. (1984). The Oxford Illustrated History of Britain. Oxford, Oxford University Press.
- Mowery, D. and N. Rosenberg (1989). Technology and the Pursuit of Economic Growth. Cambridge, Cambridge University Press.
- Mowery, D. and N. Rosenberg (1998). Paths of Innovation: Technological Change in 20th-Century America. Cambridge, Cambridge University Press.
- Mundy, J. H. (1991). Europe in the High Middle Ages 1150-1309, Folio Society.
- North, D. and B. Weingast (1989). "Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England." Journal of Economic History **XLIX**(4): 803-832.
- North, D. C. (1981). Structure and Change in Economic History. New York, W. W. Norton.
- North, D. C. (1990). Institutions, Institutional Change and Economic Performance. Cambridge, Cambridge University Press.
- North, D. C. and R. P. Thomas (1973). The Rise of the Western World: A New Economic History. Cambridge, Cambridge University Press.
- Pettifor, A. (2002). Chapter 9/11? Resolving International Debt Crises - The Jubilee Framework for International Insolvency. London, Jubilee Research @ New Economics Foundation.
- Popper, K. (1959). The Logic of Scientific Discovery. London, Hutchinson.

- Popper, K. (1961). The Poverty of Historicism. London, Routledge.
- Previts, G. and B. Merino (1998). A History of Accountancy in the United States. Columbus OH, Ohio State University Press.
- Rogers, J. S. (1995). The Early History of the Laws of Bills and Notes: A Study of the Origins of Anglo-American Commercial Law. Cambridge, Cambridge University Press.
- Schelling, T. (1980). The Strategy of Conflict. Cambridge, MA, Harvard University Press.
- Sharpe, J. A. (1997). Early Modern England: A Social History 1550-1760. London, Arnold.
- Simpson, A. W. B. (1975). A History of the Common Law of Contract: The Rise of the Action of Assumpsit. Oxford, Oxford University Press.
- Skidelsky, R. (1983). John Maynard Keynes: Hopes Betrayed 1883-1920. London, Macmillan.
- Skidelsky, R. (1992). John Maynard Keynes: The Economist as Saviour 1920-1937. London, Macmillan.
- Sloan, A. P. (1965). My Years with General Motors. London, Sidgwick and Jackson.
- Smith, A. (1776). The Wealth of Nations. London, Penguin.
- Soto, H. de (2000). The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else. London, Bantam Press.
- Stein, P. (1999). Roman Law in European History. Cambridge, Cambridge University Press.
- Teetor, P. (1962). "England's Earliest Treatise on the Law Merchant: The Essay on *Lex Mercatoria* from *The Little Red Book of Bristol* (circa 1280 AD)." American Journal of Legal History 6(6): 178-210.
- Thomas, E. (1999). Coalbrookdale and The Darbys. York, Sessions.
- Treacy, M. and F. Wiersema (1995). The Discipline of Market Leaders. Reading MA, Addison-Wesley.

Vachal, K. (1993). The Interstate Commerce Commission: Past and Present. <http://www.ndsu.nodak.edu/ndsu/ugpti/SPpdf/SP111.pdf>, North Dakota State University.

Weintraub, E. R. (2002). How Economics Became a Mathematical Science. Durham NC, Duke University Press.

Werner, W. and S. Smith (1991). Wall Street. New York NY, Columbia Univeristy Press.

Williamson, O. E. (1988). The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting. New York, Free Press.

Zuboff, S. (1989). In the Age of the Smart Machine, Basic Books.