COMMERCIAL LAW & COMMERCIAL PRACTICE

EDITED BY
SARAH WORTHINGTON

HART PUBLISHING

Commercial Law and Commercial Practice

This edited collection brings together leading scholars and practitioners from various jurisdictions with chapters and commentaries coordinated around the theme of alignments and misalignments between commercial law and commercial practice. The purpose of the book is to prompt a more critical and constructive reassessment of current commercial law and its practices, and to instigate a more fruitful dialogue between academics, judges, law reformers and practitioners.

The result is a series of provocative and challenging essays addressing an enormous range of problems that are of intimate concern to commercial practice. Some essays focus on broad themes, such as globalisation and trust. Others address more specific issues, such as contract interpretation or constraining modern management. Yet another group targets special problems, such as dematerialisation or super-priority, in order to assess the success of commercial law in meeting commercial demands. The depth and breadth of issues addressed is a credit to the authors. Taken as a whole, the volume makes some pointed suggestions for improving the practices and processes, and indeed the future progress, of commercial law.

Commercial Law and Commercial Practice

Edited by

SARAH WORTHINGTON
Professor of Law, London School of Economics and Political Science

•HART•
PUBLISHING

OXFORD AND PORTLAND OREGON
2003

Published in North America (US and Canada) by
Hart Publishing
c/o International Specialized Book Services
5804 NE Hassalo Street
Portland, Oregon
97213-3644
USA

Hart Publishing is a specialist legal publisher
based in Oxford, England. To order further copies of this book
or to request a list of other publications
please write to:

Hart Publishing, Salters Boatyard, Folly Bridge,
Abingdon Rd, Oxford, OX1 4LB Telephone: +44 (0)1865 245533
Fax: +44 (0) 1865 794882 email: mail@hartpub.co.uk
WEBSITE: http//:www.hartpub.co.uk

British Library Cataloguing in Publication Data
Data Available

ISBN 1-84113-438-4 (hardback)

Typeset by Olympus Infotech Pvt. Ltd., India, in Sabon 10/12pt
Printed and bound in Great Britain by
Biddles Ltd, *www.biddles.co.uk*

Contents

Part 3: Adapting Commercial Law to Modern Conditions

Part 4: Commercial Terms for Commercial Ends

Part 5: Controlling Modern Management

Part 6: Moving Forward: Law and Practice

Contributors

PROFESSOR WILLIAM T ALLEN, Nusbaum Professor of Law & Business, New York University, USA, Director NYU Center for Law and Business

PROFESSOR DOUGLAS G BAIRD, Harry A Bigelow Distinguished Service Professor, University of Chicago, USA

DR JOANNA BENJAMIN, Reader in Law, London School of Economics and Political Science, Member of the Bank of England's Financial Markets Law Committee, and Consultant at Clifford Chance, London, UK

WILLIAM BLAIR, QC, 3 Verulam Buildings, London, UK. Visiting Professor, London School of Economics and Political Science, UK

PROFESSOR ROBERT BRADGATE, Professor of Commercial Law, Institute for Commercial Law Studies, University of Sheffield, UK

PROFESSOR MICHAEL BRIDGE, Professor of Commercial Law and Dean of the Faculty of Laws University College London, and Director of Legal Research at Norton Rose Solicitors, UK

PROFESSOR MICHAEL BRYAN, University of Melbourne, Australia

PHILLIP CAPPER, Lovells, London, UK

PROFESSOR HUGH COLLINS, London School of Economics and Political Science, UK

ROSS CRANSTON, QC, MP, Visiting Professor, London School of Economics and Political Science, UK

PROFESSOR KEVIN E DAVIS, Associate Professor, Faculty of Law, University of Toronto.

PROFESSOR SIMON DEAKIN, Robert Monks Professor of Corporate Governance and Fellow of Peterhouse, University of Cambridge, UK.

PROFESSOR DEBORAH A DEMOTT, David F Cavers Professor of Law, Duke University School of Law, USA, and, since late 1995, the Reporter for the Restatement Third, Agency.

PROFESSOR ANTHONY J DUGGAN, Associate Dean and Iacobucci Chair, Faculty of Law, University of Toronto, Canada, and Professorial Fellow, Faculty of Law, University of Melbourne, Australia

ROD EDMUNDS, Queen Mary, University of London, UK

DAVID GOLD, Head of Litigation and Arbitration, Herbert Smith, London, UK

PROFESSOR RICHARD HOOLEY, King's College, London, UK

ADRIAN HAMILTON, QC, 7 King's Bench Walk, formerly Deputy High Court Judge (Commercial and Admiralty Courts), London, UK
PADDY IRELAND, University of Kent, Canterbury, UK
TONY KING, Head of HR Development, Clifford Chance, London, UK
GALYA LEVY, Graduate Fellow, NYU Center for Law & Business, New York University, USA
PAUL LOMAS, Freshfields Bruckhaus Deringer, London, UK
JOHN LOWRY, Centre for Commercial Law Studies, Queen Mary, University of London, UK
SIR JONATHAN MANCE, Lord Justice of the Court of Appeal of England and Wales
PROFESSOR EWAN McKENDRICK, Professor of English Private Law, University of Oxford, Fellow of Lady Margaret Hall; Barrister, 3 Verulam Buildings, Gray's Inn, London.
PROFESSOR DAVID MILMAN, CMS Cameron McKenna Professor of Corporate and Insolvency Law, Centre for Law and Business, University of Manchester, UK
DR RIZWAAN MOKAL, Lecturer in Laws, University College London; Research Associate, Centre for Business Research, Cambridge University, UK
GUY MORTON, Freshfields Bruckhaus Deringer, London, UK
SIR JOHN MUMMERY, Lord Justice of the Court of Appeal, UK
CATHERINE NEWMAN, QC, Maitland Chambers, London, UK
DR PHILIP RAWLINGS, University College London, UK
LORD JOHAN STEYN, Lord of Appeal in Ordinary, House of Lords, UK
DANIEL TAN, Lecturer in Law, Pembroke College, Oxford, UK
PROFESSOR CATHERINE WALSH, Faculty of Law, McGill University, Montreal, Canada
ANDREW WHITTAKER, General Counsel to the Board, Financial Services Authority, UK
PHILIP WOOD, Allen & Overy, Visiting Fellow at Oxford, Cambridge, QM and LSE, University of London, UK
PROFESSOR SARAH WORTHINGTON, London School of Economics and Political Science, UK
NIK YEO, Barrister, Fountain Court Chambers, Temple, London, UK
PROFESSOR JACOB ZIEGEL, Professor of Law Emeritus, University of Toronto, Canada

Introduction

Aligning Commercial Law and Commercial Practice

SARAH WORTHINGTON

THIS COLLECTION OF essays is intended to spark debate. The goal is to promote a better dialogue between academics, judges, law reformers and practitioners, and to initiate a more constructive reassessment of current commercial law and its practices. Both aspects are crucial if the underlying debates are to have a noticeable and worthwhile impact on legal developments. Early drafts of these essays were first discussed at a two-day round-table seminar held at the LSE in November 2002. That early face-to-face debate was merely a preliminary step in exploring these issues. The published collection will take the ideas to a wider audience. This strategy is becoming increasingly common, although the breadth of the agenda behind this collection is novel.

I. RECOGNISING NORMS IN ENGLISH COMMERCIAL LAW

All of the essays and commentaries in this collection are coordinated around the theme of alignments and misalignments between commercial law and commercial practice. The time is ripe for critical reflection amongst commercial lawyers. It seems obvious that commercial law—the law of the merchants—should, by its very nature, be constantly under review. Its function is to facilitate and, when necessary, constrain, the commercial activities of businesses and entrepreneurs. As business and business practices evolve, and as social expectations change, so too must the relevant legal rules. Yet examination of English commercial law exposes little more than incremental and piecemeal change over the past century. The enormous strides made over a century ago by both judges and legislatures have given way to tinkering at the margins. Worse still, even the tinkering is largely driven by EU challenges and company and financial services Directives rather than by serious internal domestic reassessment. The product of this approach is inevitably patchy and in many places outdated. The genius of English commercial law is losing out to its more aggressive and focused international competitors. To allow this to continue seems at best short-sighted and at worst cowardly, given the enormous potential still locked away in English commercial legal practices.

The way forward is not easy, however. Judges can only deal with the issues presented to them. Their response is necessarily issue-specific, party-oriented and rule-based. Nonetheless, especially in the House of Lords, there is an increasing willingness to consider explicitly the policy ramifications of the specific matters that do fortuitously reach the court for consideration. *Westdeutsche Landesbank Girozentrale v Islington London Borough Council*[1] and *Barclays Bank plc v O'Brien*[2] are perhaps two of the most obvious cases from the last decade. Legislatures and law reform bodies are not so constrained in their approach, and yet all too often these institutions are content to concentrate on the particular, delving into the detail of troublesome but narrowly focused legal rules or lacunae in the rules, before ever asking the important questions about more general overarching aims and objectives. The past decade of company law reform, for all its worth, might be criticised on this basis. The DTI promised radical overhaul,[3] but the earlier more narrowly focused work of the Law Commission on shareholder remedies and directors' duties[4] provided pre-established pathways that appear to have seriously compromised potentially more imaginative approaches.[5] Professor Jacob Ziegel touches on some of these issues (and many others) in chapter 22.

What should be done? The micromanagement that currently exists is undoubtedly necessary, but it is not designed to deliver coherent, comprehensive modern law reform. For that, it is necessary to stand back and survey the entire field. The common law judicial system is not designed to deliver this form of oversight. Nor, indeed, is the legislature: the government's political will cannot embrace the project. In Britain, there is no recognised formal institution charged with the task and empowered to implement, or even promote, its findings. The Law Commission's role is much more tightly regulated, and the government seems philosophically disinclined to change this.[6] The task is therefore left largely to lobby groups, including academic lawyers. A united front might prove far more effective, with academic lawyers, practitioners, judges and various other

[1] [1996] AC 669.
[2] [1994] 1 AC 180.
[3] DTI, *Modern Company Law for a Competitive Economy* (1998).
[4] Law Commission, *Shareholder Remedies* (1997, Law Com No 246); Law Commission, *Company Directors: Regulating Conflicts of Interest and Formulating a Statement of Duties* (1999, Law Com No 261, Scot Law Com No 173).
[5] See S Worthington, 'Reforming Directors' Duties' (2001) 64 MLR 439–58.
[6] See, for example, the government White Paper, *Modernising Company Law* (2002, Cm 5553–I), pp 48–9, rejecting the need for a formal institution to maintain general oversight of English company law.

parties all making use of the tools at their disposal to analyse the problems and advance potential solutions. In this vein, much might be learnt from the role played in the United States by the American Law Institute.

The first task of such a coordinated body should be to settle the strategic direction. What is the function of commercial law, and what overarching objectives should English commercial law pursue? Some sort of agreed position on this is essential if the more specific issues are to be addressed in any meaningful way. Consensus is unlikely to be easy, yet some agreement is crucial to determining the form and function of commercial law. Many of the essays in this collection confront this issue, some directly, some indirectly. See especially Sir John Mummery (chapter 2), Paddy Ireland (chapter 4) and Professor Robert Bradgate (chapter 23).[7] The diversity in their approaches is predictable, confirming, if that were needed, the urgency of taking coordinated steps to identify a coherent and defensible strategy for advancing the future direction of the law.

II. IDENTIFYING SHORTCOMINGS IN MODERN COMMERCIAL LAW

Once the broad goals of English commercial law have been settled, it should be possible to pinpoint those legal rules that seem especially effective or especially ineffective. Several strategies could be usefully employed. Quantitative and qualitative investigations may well indicate how commercial parties themselves view the successes and shortcomings of the English jurisdiction. Traditional legal scholarship should also prove fruitful: cross-border comparisons and focused engagement with international legal and political debates is likely to lead to important insights.

Notice, however, that the outcomes of this second stage evaluation are critically dependent upon the precise standards against which the legal rules are being judged. It follows that this second step cannot be undertaken in any meaningful way until the first step has been completed successfully—although the common practice is often otherwise. More importantly, the first step must identify norms with sufficient particularity to enable sensible discrimination. The touchstone may well be Lord Steyn's suggestion that the law should uphold 'the reasonable expectations of honest men.'[8] But this is too vague to provide a useful normative yardstick. Reasonable expectations vary. For example, many would regard the genius of English commercial law as its flexibility and liberal approach to freedom of contract. These particular features may well be responsible for

[7] Also see, for example, M Bridge, 'The Future of English Private Transactional Law' (2002) 55 CLP 191, and the references cited.
[8] (1997) 113 LQR 433.

delivering several crucial commercial developments, such as negotiability, the trust, and the charge. Judged against these free-market standards, many of the legal rules that currently regulate commercial agreements could be seen as overly paternalistic. If a mismatch is confirmed, then the offending rules might usefully be eliminated, or at least modified, in order to facilitate a return to the recognised core values of English commercial law, with their emphasis on party autonomy, ease of contracting, and certainty in risk management. More recently, however, other normative approaches are proving influential. According to these equally 'reasonable expectations', the parties' bilateral agreement is not necessarily everything: third party interests may also need protection. Equity's imposition of fiduciary duties and duties of confidence provide early and influential illustrations of the way the law has addressed these wider issues. It remains true, however, that western democracies are clearly committed to individual freedom and autonomy, so any legal regulation should be consciously limited to the minimum necessary to achieve the identified social goals.

Some of the inadequacies identified at this second stage will reflect deep-seated philosophical disjunctions between agreed modern norms and entrenched legal rules. The restrictive rules relating to variation of contracts, or to agreed remedies clauses, are often regarded as inconsistent with party autonomy. Other disjunctions are different, however. They relate to unanticipated differences between old and new commercial practices. For example, much of our orthodox commercial law was designed to deal with directly negotiated, short-term, geographically limited arrangements relating to tangible property. Now, however, the reality is that many agreements are agency-mediated, long-term, international negotiations relating to intangible assets. Modern commercial law must therefore address the practical problems of fiduciary management, globalisation, conflict resolution (both between parties and between national legal systems), and asset protection. Predictably, many of the essays in this collection address these apparently escalating problems, some commentators agreeing that there are crucial shortcomings which need remedying, others arguing that there is nothing terribly new in all of this—as Professor Michael Bridge pithily observes, a lot of e-law turns out to be old law in a new setting (chapter 9). See especially the essays by Professor Ross Cranston (chapter 1), Dr Joanna Benjamin (chapter 10), Professor Catherine Walsh (chapter 16), Professor Michael Bryan (chapter 18) and Professor Douglas Baird (chapter 20).

III. IMPLEMENTING APPROPRIATE CHANGES IN THE LAW

The third step in negotiating the path of modern law reform is obvious. Once the disjunctions between current rules and agreed norms have been identified, then strategies for implementing appropriate legal change need to be developed. This demands some rare talents. Pertinent, coherent outcomes are unlikely unless the reformer can take a step back from the detailed doctrinal rules, keep an eye on the agreed overarching goals, and develop principled strategies to achieve the desired ends.

Adequate reform cannot simply focus on doctrine, even if this is where the identified disjunctions lie. For a start, reforms generated in this way are unlikely to involve more than tinkering at the margins. A doctrinal approach inevitably confines thinking within legal categories that may no longer be apt. Consider, for example, the increasingly blurred boundaries between debt and equity, or property and obligation. More subtly, and perhaps more dangerously, a doctrinal focus encourages functional assumptions that may misrepresent commercial practice. The possibility is illustrated within this collection: in chapter 17, Dr Riz Mokal argues that the real function of the floating charge is not to provide better security for debts, but better control over the debtor. The difference matters, if subsequent reform is to make any sense. A focus on principle rather than doctrine, and function rather than form, would prompt far more radical and far-reaching revisions.

Even these crucial oversights and insights cannot solve all the problems, however. Difficult choices must still be made between different reform options. As Professor Tony Duggan argues in chapter 21, it is important that these choices be made carefully, thoughtfully and indeed proactively: we must not simply 'muddle through'. There are costs and benefits associated with every choice, and the legal regime's internal coherence to be considered. Whatever social or economic factors we decide ought to be weighed into the balance, Duggan suggests that the analytical framework of law and economics can assist in resolving the competing claims. He argues powerfully for this sort of rigorous, focused approach from our judges. We might expect the legislature and advisory bodies to be even more aggressive on this front.

None of this requires that these carefully chosen reform strategies be startlingly original. The old tools may prove perfectly adequate, even if the old doctrinal classifications are not. Orthodox compensation principles, or fiduciary rules and equitable proprietary interests, for example, may provide quite adequate routes for imaginative intervention. All the same, there is much that might be learned from the different approaches adopted elsewhere, whether in common law or civilian jurisdictions. The law of securities provides a ready illustration. The radical reforms introduced in the United States by the Uniform Commercial Code, Article 9, have now been revised and adopted in many other jurisdictions, although not yet in the United Kingdom.[9] Adoption does not imply replication, however.

[9] In the UK, the issue has been the subject of several unimplemented recommendations over the past thirty years (see Ziegel, ch 23 below). Perhaps the current work of the Law Commission is set to prove more successful: Law Commission, *Registration of Security Interests: Company Charges and Property Other Than Land* (2002, Consultation Paper No 164).

National differences work to ensure that even when the rules are formally identical, the results are not. In any event, as Professor Walsh points out in chapter 16, there is often surprisingly little consensus on the most advantageous form of the rules themselves. Indeed, whatever the form, any statutory statement brings with it its own peculiar difficulties and compromises: Professor DeMott explores some of the relevant issues in chapter 3. Moreover, not all national divergences are codified, of course. Consider the radically different rules adopted by different legal systems to protect property owners from unauthorised dispositions of their property. The uncodified English rules are very protective; the codified civilian ones are not. Again, the comparisons are instructive, especially given current English concerns about the merits of such seemingly ruthless treatment of the losing parties. As with all these comparisons, the goal is not necessarily harmonisation, but simply selection of the best possible approach.

Useful general rules to guide these reform practices cannot be mandated at the outset. Reforms, whether judicial or legislative, must range widely. They must address both party-party and public issues, and must embrace the protection of interests and values as well as rights. Deliberate efforts must be made to maintain both vertical and horizontal coherence within the legal system. All systems evolve, so some type of reform is inevitable regardless of the national strategy. The crucial issue is how reform is managed. Even seemingly stable issues, such as basic contract interpretation, are currently generating enormous debate: see the four quite different contributions by Lord Steyn, Professors McKendrick, Allen and Collins, and Galya Levy in Part 2 of this collection. These particular debates are revealing. They suggest a growing inclination to escape the orthodox contract analytical straightjacket of offer, acceptance and consideration, and its defined moment of commitment to the deal. Instead, the modern goal appears to be a far more context-sensitive approach to the agreement. In function, if not in form, the parallels with civilian notions of good faith are clearly evident. On other fronts, activity on the public or social policy side of the reform agenda is no less vibrant. We might think that most economic and social problems are already well served by the likes of the common law rules relating to fiduciaries and breach of confidence, or the statutory rules on consumer protection and market abuse, for example. This is not necessarily so: notice the fierce and very public ongoing debates surrounding issues of corporate governance. All of this merely serves to emphasise the importance of maintaining constant formal oversight of the legal system; muddling through will not do.[10]

To reiterate, this third step in the reform process requires much more than simple, issue-specific reaction to a problematic rule. If our national system

[10] See Part I of this chapter, above.

of commercial law is to remain attractive to all affected parties, then any reform must play a calculated role in delivering a coherent, sophisticated, discriminating regime which purposefully advances certain public and widely agreed overarching norms. Such a system will have inbuilt horizontal coherence, as well as the more familiar and easily delivered vertical coherence.

IV. COMMERCIAL LAW ON THE INTERNATIONAL STAGE

If these three steps were implemented immediately, we would eventually have a domestic commercial law system we could be proud of. The changes would not happen overnight: the decay has been too long in the making. Before going down that route, however, there is one further concern that might influence the strategy we choose to adopt. Commercial law is obviously becoming increasingly international, even if that trend began with the Greeks. Internationalisation creates its own problems. This is especially so if English commercial law evolves down one path while other jurisdictions proceed down another. We will then need well-developed conflict of laws rules to address the inevitable problems of cross-border commercial transactions: see the issues advanced by Nik Yeo and Daniel Tan (chapter 14). Moreover, the likelihood of such separate and distinctive evolutionary paths increases in direct proportion to the difficulty we have in explaining our rules to others. The dilemma is clear. We value the evolutionary flexibility inherent in the common law system: its commitment to human liberty and minimal intervention combined with its decentralised system of judge-made laws undoubtedly gives rise to a great variety of rules and responses to relatively similar problems. Only the strongest are likely to survive. These may well be the 'best' rules, but—persisting with the Darwinian analogy—they are unlikely to flourish elsewhere unless they are readily adaptable to different terrains. Civilian codes have this advantage. Indeed, much of the world is now governed by codes that still exhibit traces of their ancestral Roman legal traditions. The message is clear. If we want our commercial law expertise to extend beyond our own shores, then we need not only the best rules, but also the most easily transplantable ones. Again, the agenda is not necessarily harmonisation, but simply communication. This does not mean that a 'commercial code' is necessarily the best choice,[11] even if it is the most obvious one.

[11] But see the arguments in RM Goode, 'Insularity or Leadership? The Role of the United Kingdom in the Harmonization of Commercial Law' (2000) 50 *International and Comparative Law Quarterly* 751, at 761–63.

Along the spectrum between an explicit code and a purely precedent-based system with the rules hidden in a myriad of cases, there are several other attractive alternatives. Again, the models can be found elsewhere. UNIDROIT and UNCITRAL, the American Law Institute and, increasingly, European-based higher-level groups have all produced a range of alternative strategies. We must at least be aware of the competition, whatever we choose to do about it.

V. THIS COLLECTION

The views outlined in the preceding few pages are personal, although there is much in this collection that might be used to support the suggested reform agenda. Nevertheless, the chapters in this book are not designed to support just one particular message. The agenda underpinning this book is much broader, and the organisation of its contents is therefore different. The goal is simply to prompt debate. Given that, it is irrelevant that different chapters within the collection advance conflicting views. There is no agreed 'party line'. Instead, the chapters are divided into six Parts, with each Part collecting together papers that are loosely coordinated around certain broad themes with no thought for a preconceived reform agenda.

Essayists were initially given no more guidance on subject-matter than a bald request to reflect on specific areas where commercial law fails to match commercial expectations, and areas where it succeeds, and then to provide a provocative analysis of the reasons for this. Notwithstanding the open-ended invitation, certain themes are clearly evident in the end product. There are twenty-three essays in all, eleven with published commentaries. The seminar which preceded this volume was run under the 'Finn rules': papers were distributed in advance; commentators spoke briefly to a particular paper; authors were given a limited opportunity to respond; and the floor was then opened for discussion. Essays by academics received comment by practitioners, and vice versa. Senior members of the judiciary also participated, providing both essays and commentary. All three groups acted as chairs for various sessions.

In organising the seminar and producing this volume, I have incurred large debts to several groups of people. First and foremost, of course, I am enormously grateful to all the authors, commentators and chairs who gave so enthusiastically of their time and efforts. The success of the seminar itself and the relatively painless delivery of this large volume are testament to their commitment and good humour. The two-day seminar was generously funded through the Annual Seminar Competitions run by both the Society of Legal Scholars and the Modern Law Review, with further funding from the LSE Law Department, and additional sponsorship from Nabarro Nathanson. Equally importantly, four rather special law students worked

tirelessly behind the scenes to ensure a smooth production: Katherine Worthington gave me superb organisational assistance in the months leading up to the seminar; she, Madeleine Wanner and Carrie Suen also provided intensive support during the seminar proceedings themselves and in the days immediately preceding the event; and Tim Akkouh contributed much-needed proof-reading and editorial assistance. Finally, Richard Hart has yet again managed the publication of this volume with his now legendary professionalism, efficiency and good humour. Thank you to all concerned.

The collection is divided as follows. Part I comprises chapters on the issues that most commercial lawyers regard as providing pressure for change. There is discussion of the political and economic issues associated with globalisation, issues of good faith, the growing trends towards codification, and the increasing perception that government should regulate individuals to promote the greater common good. Part II is devoted to the interpretation of contract terms. Practising lawyers regard this issue as fundamental, and under-analysed by both the courts and academics. Part III then turns to the practical issues of commercial context, and considers how various factors do, and should, play into the development of a more satisfactory modern commercial law regime. Part IV continues this theme in a different vein, examining various strategies that the parties themselves have adopted to deal with modern circumstances. An indicative cross-section is all that is possible in the space allowed. Nevertheless, different authors focus on matters as diverse as cross-border problems, large-scale employment issues, and the management of insolvency risks. Part V is devoted entirely to modern problems of constraining management. Finally, and by way of conclusion, Part VI contains three quite different chapters all directed at assessing how we might better manage legal developments.

A final word seems apt. Many people insist that the study of law needs to be interdisciplinary if it is to have an impact. Right now, however, the best 'interdisciplinary' collaboration might well be between academics, practitioners and judges. Dialogue across the professional boundaries is crucial. If this volume contributes to that process, then the efforts of all those involved will have been worthwhile.

Part 1

General Pressures for Change

1

Globalisation: Its Historical Context

ROSS CRANSTON, QC MP

THIS CHAPTER IS part of a larger project to examine the historical background to Anglo-American commercial law.[1] Its particular focus is international capital flows in the nineteenth and the very first part of the twentieth century. It takes as a starting point some of the massive literature on globalisation, and in the process seeks to pour cold water on some of the fashionable talk. It is very much work in progress. The focus at this point is on London, rather than New York. The emerging theme, however, is that the ground was laid in the nineteenth century for many of the legal institutions and techniques used in modern day international lending.

Globalisation as an organising concept for academic work is fraught with difficulty. Although the word was apparently coined in the 1940s, it was not until the 1980s that it acquired a real significance. It has obvious implications for law studies—that they should become more interdisciplinary, comparative and international. The difficulty is in determining what the concept encompasses. For some like Robert Reich, it is about the fragmentation of world production, the expansion of world trade or the way market liberalisation has swept the board internationally as a basis for public policy.[2] For sociologists such as Harvey, it is about the homogenisation worldwide of consumer culture and that relations between people now transcend territorial space.[3] Those concerned with international relations may take globalisation as a direction to focus on international organisations such as the World Trade Organisation, the International Monetary Fund and the World Bank, or possibly at the present time the internationalisation of terrorism. For theoreticians like Giddens, globalisation is about 'the disembedding of social systems' or 'the lifting out of social relations from local contexts of interaction and their restructuring across indefinite spans of time—space'.[4]

[1] R Cranston, 'Doctrine and Practice in Commercial Law', in K Hawkins (ed) *The Human Face of Law* (1997).
[2] R Reich, *The Work of Nations. Preparing Ourselves for 21st Century Capitalism* (1991).
[3] D Harvey, *The Condition of Postmodernity* (1990).
[4] A Giddens, *The Consequences of Modernity* (1990), p 21.

Whatever view ones takes of globalisation, the key point for this paper is that it is not a new phenomenon. Capital flows are one indication of globalisation.[5] In 2001, the Bank for International Settlements reported in its *Quarterly Review* on the cross-border claims on BIS reporting banks—it was $617.2 billion in loans and $233.7 billion in securities. (Foreign direct investment by multinational companies is a separate topic, not addressed in these figures or in this chapter.) Yet, as we will see in a moment, the flows in the nineteenth century were comparable, although the strict division between loans and securities was not as clear-cut. Thus, on globalisation, I end up with William Twining: '[I]f one blows away some of the froth of the highly repetitive literature, one given is that globalisation, far from being a new phenomenon, has a history that stretches back at least two centuries'.[6]

Generally, there is little writing on the legal context of international lending although there is some on the role of law firms and associated actors.[7] The institutions and techniques for international syndicated lending and securities issues used in the present day are dealt with in a number of standard works.[8] The argument in this paper is that we can see these techniques and institutions already developing in the nineteenth century. This is not an argument that all the institutions and techniques were full blown, but the surprise for the present day lawyer is that so many of them were already there, developed in the main by the practice of investment banks, but assisted by their legal advisers. There is no existing account of this history. Apart from the huge volume of work by economic historians, my starting point has been with the history of leading investment banks and with the small number of histories of leading law firms.[9] I have obtained access to the archives of three London based banks and four law firms.[10] Clearly there is also occasional litigation and nineteenth century legal writing. What emerges from my study at this stage is as follows.

I. GLOBAL CAPITAL IN THE NINETEENTH CENTURY

The history of international lending in the years 1800–1914 has been told many times. The title of a classic study, Feis' *Europe: The World's Banker 1870–1914* (1930), indicates how until World War I the flow was from

[5] See eg, E Helleiner, *States and the Reemergence of Global Finance* (1994).

[6] W Twining, 'Reviving general jurisprudence', in M Likoksky (ed), *Transitional Legal Processes* (2002), p 3.

[7] Eg, J Flood, 'Capital Markets, Globalisation and Global elites', in *ibid*.

[8] Eg, R Cranston, *Principles of Banking Law* 2nd edn (2002).

[9] Bank histories are mentioned throughout. On law firms see eg, J Slinn, *Clifford Chance* (1993); A Dean, *William Nelson Cromwell* (1957).

[10] Notably The Rothschild Archives [hereafter RA]; Barings' Archives [hereafter BA]; National Westminster Bank Archives [hereafter NWA]. I am greatly indebted to the respective archivists for their assistance.

Britain and, later, countries such as Germany, to other parts of the world—North and South America, Central and Eastern Europe and elsewhere. It was mainly for governments and railways. Europe's position as banker to the world came to an end with the First World War. Loans floated in the United States on behalf of the British and French governments extinguished the position of the United States as a net debtor and changed American perceptions of foreign lending. So from 1920–1930, the United States became a world banker: it lent $10 billion abroad, 40 per cent as foreign direct investment, half to Europe, just over a quarter to South America.[11]

The techniques for international lending had been developed in Amsterdam in the eighteenth century, by banks such as Hope & Co.[12] But in the nineteenth century the focus of international lending moved from Amsterdam to London. Those great investment banks of the nineteenth century, the Rothschilds and Barings, became loan brokers for the legitimist regimes of Europe after 1815, using the Congress of Europe meetings as their auction rooms.[13] Other investment banks followed. After some disastrous investments in Greece and South America in 1820s, the banks turned to North America, first, to state government paper.[14] Then in the 1840s southern American states defaulted and US state government securities consequently lost their allure to foreign investors, who turned increasingly to railroads.[15] The demand for capital to build a national rail network in so geographically large a country was enormous. The response through foreign investment in 'American Rails' was massive, mainly in bonds rather than shares. There were numerous public issues in London, but a snapshot in 1886 also uncovered over 100 American issues not listed on the stock exchange yet 'known' in London.[16]

There were issues for American Rails on the continent as well. Indeed it is well to remember that both France and Germany were substantial providers of capital to the world in the nineteenth century. However, the bulk of German investment, for example, was in other European countries such as the Austro-Hungarian Empire and the Balkans, whereas British investment was more widely dispersed in North America and the Empire. Both Britain and Germany held roughly proportional investments in South America. By 1914, Britain was exporting more capital than Germany and France together.

[11] B Eichengreen 'The US capital market and Foreign Lending, 1920–1955; in J Sachs (ed), *The Developing Country Debt and the World Economy* (1989).
[12] L Neal, *The Rise of Financial Capitalism* (1990).
[13] P Mathias, *The First Industrial Nation*, 2nd edn (1983), p 294.
[14] R Sylla, J Wilson, R Wright, 'Trans-Atlantic Capital Market Integration, 1790–1845', unpublished paper, IMF, 2002.
[15] See eg, M Edelstein, *Overseas Investment in the Age of High Imperialism: the United Kingdom 1856–1914* (1982); M Wilkins, *The History of Foreign Investment in the United States to 1914* (1989); C Kindleberger, *A Financial History of Western Europe* (1984).
[16] D Adler, *British Investment in American Railways 1834–1898* (1970).

Britain by the late nineteenth century had extremely efficient capital markets for issuing and trading first class securities. The advantage Germany had domestically, however, was that through its system of bank lending it was able to allocate capital to industrial development and innovation. Britain has been unable to emulate that to the present day despite being 'banker to the world'. Indeed at its peak the British share of total global investment was almost 80 per cent, far above what the United States ever achieved when it was at its zenith in the twentieth century, which was about 50 per cent in 1960.

There has been considerable debate about which investors were supporting foreign lending. In a useful corrective, Platt argues that simply because a loan was issued in London did not necessarily mean it was taken by British investors. He demonstrates that paper issued in London often flowed back to the domestic market where the loan was raised.[17] Certainly it cannot be assumed that quotation on the Stock Exchange in London means that English investors supported the issue, nor did the face value of a loan necessarily have a close relationship to the capital actually invested. However, by the turn of the century there was no doubt that British capital transfers to other countries—both for governments and investments such as railways and mines—was enormous. By 1913 foreign investments were 9.1 per cent of gross national product, a level not again reached until the 1990s. Ferguson sums up the position thus: 'The share of British wealth invested abroad rose from 17 per cent in 1870 to 33 per cent in 1913. No other country came close to this level of foreign investment: the closest, France, had foreign assets worth less than half the British total, Germany just over a quarter. Britain accounted for something like 44 per cent of all foreign investment on the eve of the First World War'.[18] The First World War ended all this. Britain became a bit player and the United States emerges in the 1920s as the great international lender.

II. BANKS AND BORROWERS

> Those [Rothschild and Baring], and the truly liberal Laffitte,
> Are the true lords of Europe. Every loan
> Is not a merely speculative hit, but seats a nation or upsets a throne.
>
> (Byron, *Don Juan*, Canto XII)

Let us turn now to some of the more legal aspects of international lending in the nineteenth century, beginning with the relationship between borrowers

[17] D Platt, 'British Portfolio Investment Overseas before 1870: Some Doubts' (1980) 33 *Econ History R* 1.

[18] N Ferguson, *The Cash Nexus. Money and Power in the Modern World*, 1700–2000 (2001), p 297 (footnotes omitted).

on the one hand and banks on the other. Compared with the present day some aspects seem to have been very informal. These days borrowers might ask banks to tender for a particular loan, and 'beauty parades' have become common. There was tendering for loans in the nineteenth century, but it seemed to be just a matter of banks putting forward their proposed terms. Grellier, in his book published in 1812, *The Terms of All the Loans …*, recognises that there were biddings for UK Government loans in the early nineteenth century.[19] Barings was disconcerted when the Russian Government proposed to offer a contract of loan in 1858 to other tenderers. Bids were put in and ultimately the contract was awarded to Thomson Bonar, a relatively minor firm.[20] Tendering, however, does not seem to have become a matter of law and issuers exercised a wide discretion as to who would get a loan contract. No doubt they would have their eye on the future as well, because to offend a Barings could well store up problems when loans were needed in harder times.

An issue early recognised as having legal implications, which still surfaces, is that negotiations may be entered, and agreement reached but some time elapses before a formal contract is signed. To what extent is a party bound before the formal contract is signed? From the point of view of a bank, tying down the borrower may be crucial if there has been competition from other banks. Conversely, the borrower may desire the commitment of the bank, if the terms are favourable, especially if the market might turn for the worse before the formal contract is completed. In the 19th century, agents in London would have authority to agree terms, so there was not always a need to refer matters back to the borrower's country. This was important before the telegraph was introduced in the 1840s and underwater cable from the 1860s. However, there was still delay in drawing up a formal contract after negotiations had been complete. In 1824 Nathan Rothschild negotiated a Brazilian Government loan for £1,686,000 at 5 per cent. Rothschild wrote to the Brazilian Ambassador with an acknowledgment of receipt clause:

> We have given directions to prepare and submit to you a formal Contract and until that shall have been executed, your Letter and this our Answer are to be considered as constituting the Contract.[21]

It was a canny appreciation of the legal difficulties associated with the negotiations leading up to a first contract.

[19] J Grellier, *The Terms of all the Loans which have been raised for the Public Service*, 3rd edn (1812).
[20] P Ziegler, *The Sixth Great Power. Barings 1762–1929* (1988), p 175.
[21] RA 001/401A.

A. The Agreement

Perhaps the agreement which provided the template for nineteenth century lending was that resulting from the negotiation by the banker Walter Boyd of a loan for Austria in 1795. The terms of the agreement were published in what subsequently became known as *Hansard*, and therefore was widely available.[22] An examination of this agreement demonstrates the falsity of the claim that it was Nathan Rothschild who introduced payment of interest in London, instead of in the borrower's jurisdiction, and was the first to fix payment in sterling to avoid fluctuations in the exchange rate.[23]

However, for our purposes let us take an agreement which Nathan Rothschild negotiated in 1824 for a loan of £2.5 million with the kingdom of the Two Sicilies.[24] The agreement was straightforward. The parties were the Minister of Finance on the one side—authorised by the King and in accordance with law—and Nathan Rothschild 'a partner and representative' of the Rothschild houses around Europe on the other. As to payment, the Minister in the name of the government agreed to pay interest at the specified rate. To this end, the King undertook in the agreement to issue a general bond for repayment of the interest in both London and Naples. For payment of capital, the general bond was also to contain an undertaking to create a sinking fund by specified annual payment to the Rothschilds. The agreement required the general bond to be deposited in the Bank of England with certain formalities—no doubt designed to enforce the commitment to pay—and kept there for the duration of the loan.

On the basis of the general bond, definitive bearer bonds of £100 were to be issued with dividend warrants attached, and each incorporating all the provisions of the general bond. The dividend warrants were payable half yearly in London and Italy. Capital was to be paid off gradually over the period of the loan by the half yearly redemption of bonds, either by purchase in the market, or if above par by lot, the numbers chosen being published in the *London Gazette* and in Italy. Redeemed bonds were to be cancelled, no doubt to prevent fraud: the formalities for cancellation are set out in some detail, basically the bonds being cut in two, one half being placed in the Bank of England with the general bond, the other being sent to Italy. Redemption meant of course that a bondholder was no longer entitled to interest. There is correspondence in the Rothschild archives of a bondholder of one of the Brazilian loans issued by the bank presenting a summons to the Brazilian Ambassador for payment. He claims he only looked at his bond and did not see the advertisement in the *London Gazette*

[22] *Parliamentary History of England*, v 31, cols 1561–2. On the background see K Helleiner, *The Imperial Loans* (1965).
[23] S Mace, 'The Archives of the London Merchant Bank of N M Rothschild & Sons', *Business Archives*, No 64, November 1992, p 6.
[24] RA 000/401 A.

that his bond had been redeemed. Clearly the claim had no merit, since the bond itself contained the provisions about redemption.[25]

The approach in the Italian loan was typical for international lending, not simply for the Rothschilds but for other banks as well.[26] There might be variations in format. For example, an Ottoman loan of 1855, negotiated by the Rothschilds, had an agreement running just over a page, with the particulars of the proposed loan attached. That loan was also a variant on the basic approach because payment would not be made by the Rothschilds, but by the Bank of England under a separate agreement negotiated simultaneously with the main agreement.[27] In the main, however, there was little change in substance as far as nineteenth century loan agreements were concerned, at least for government issues. Bonds were to be issued, interest on them was payable in London and possibly in the borrower's jurisdiction, and a sinking fund was to be established to repay the capital by periodic redemption of the bonds. Embellishments designed to assure investors had little legal effect. Thus the Imperial Turkish loan agreement for £5 million, negotiated by Dent, Palmer & Co. in 1858, constituted a commission of five persons at Constantinople, chaired by the Minister of Finance or his nominee, with two representative of the Sultan and two appointed by the bank. The commission was the conduit of funds for London, but had no power to appropriate revenue streams such as the customs duty to be collected at Constantinople, which was supposedly assigned for payment of interest and principal.

Separate from the loan agreement itself might be agreements relating to the mechanics of the loan. Thus there might be an agreement as to the borrower opening a bank account, if this did not already exist, to facilitate payments of interest and the redemption of the bonds. Of course the profits to be made by a bank from these banking functions were often more important than those from the issue of the loan itself, which a bank might enter as a loss making activity.[28] In some cases there were separate agreements, kept confidential, for payment of additional commission to the bank for negotiating the loan.

B. Key Terms of the Agreement

The motivation behind a government loan might be obvious—to fund a war, to overcome a budgetary crisis, or even to pay interest on a previous loan. As a matter of law, however, many government loans were for the general

[25] RA 000/401B, Dispute with Mr Hodgson; Re Bond 372 of Brazilian Loan £800,000, 1829 (1857).

[26] Thus there is little difference half a century later with the Rothschilds—United States Loan, 1874, RA 000/401D. On Barings: see R Hidy, *The House of Baring in American Trade and Finance* (1949), pp 33, 60.

[27] RA 000/401B. On the background: N Ferguson, *The World's Banker* (1998), pp 584–5.

[28] K Burk, *Morgan Grenfell 1838–1988* (1989), p 32.

purposes of the state and no specific purposes would be mentioned in the agreement, prospectus or bonds. If a purpose was specified, there was little to prevent a government using the money for other purposes. Lawyers would have taken little comfort from the provision in the decree of the Russian Tsar in 1850 relating to a £5.5 million loan for completion of the Moscow-St. Petersburg railway, arranged through its banker Baron Stieglitz, and floated in London by Barings—that the account of monies derived from the loan were to be kept distinct from other state revenues and expenditure to guarantee that they were correctly allocated.[29]

Corporate lending was much more specific in terms of the purpose—for example, to build a railway or to develop a mine. Even here, however, it was difficult to see what a lender would do were the purpose clause to be breached, even if the bank still had a direct interest in enforcing the purpose clause, because it still had some of the issue. In practice, the clause might be so vague as to be impossible to say that it had been breached. There are examples of staged payments from an issue, depending on successful completion of earlier stages in the project, and this was possibly the most effective means of enforcement of a purpose clause.[30]

The conditions precedent and representations and warranties of nineteenth century lending were rudimentary, and it is here that twentieth century loans differ most from their ancestors. Certainly in terms of both government and corporate lending the legislative authority authorising the government loan or the project to be financed (eg, the building of a railway) would be set out in the loan agreement and referred to in the prospectus and bonds. So in practice the banks were treating this as a condition precedent. The need to have a loan authorised by law became more formalised with time, so that in an agreement between the Rothschilds and the Republic of South Africa for a loan issue of 5 per cent government bonds in 1892, the agreement requires that the government shall produce 'forthwith' the legal authorisation in favour of the issue and also a legalised copy of the loan.[31] Moreover, where a government loan was pursuant to an international treaty, we find provisions such as that in a Rothschild loan to the Egyptian Government in 1885 which provides as follows:

> This contract shall be construed with reference to the said Convention and Decree and any clause or provision herein contained which may be deemed to be inconsistent with the said Convention and Decree shall to that extent of such inconsistency be void and to no effect.[32]

[29] Art 5 of the Ukase of the Czar, dated 9 December 1849 (old style), BA.
[30] Loan of £1,700,000 by Barings/Hope to Moscow-Kursk Railway Company, 1871, BA, Safe Document 454.
[31] RA 000/401E.
[32] The Convention referred to was the Egyptian Financial Convention of 18 March 1885, ratified by Britain, France, Germany, Austria, Italy,Russia and Turkey. See generally D Landes, *Bankers and Pashas* (1958).

In addition to this general legal authorisation there would need to be specific authority for particular acts under a loan. For example there is a draft of a letter from the Rothschilds to the Brazilian Ambassador relating to a Brazilian Government loan in 1841 as follows: 'We have been legally advised that these bonds cannot be issued to the public, until the same have been ratified by the Brazilian Government'.[33]

As to representations, one finds few in the government loan agreements, although there might be a term not to undertake any further borrowing, for example, as long as the bank has an option to take further bonds in that issue. Such undertaking would not necessarily be on the face of the agreement, but might be by side letter. Thus when the UK Government needed to borrow in 1854 as a result of the Crimean War, Gladstone as Chancellor wrote to the Governor of the Bank of England undertaking not to issue during that Parliamentary session Exchequer bonds payable in 1858 beyond £2m. This was hardly a restraint, since the following year the Government went back to the market.[34]

British banks rarely inserted a tying clause in their loan agreements, whereby the borrower was obliged to place part of the loan to orders with manufacturers in the bank's country. The reluctance of the British Government to intervene in the direction and control of investment, and the division between financial and manufacturing interests in Britain, contrasted with the approach in France and Germany where tied loans were more common.[35] Nor would we find provisions which must have operated domestically in the United States, whereby bankers were represented on the boards of companies, particularly the railways.[36]

A key issue between bank and borrower was the obligation of the bank in relation to the securities backing the loan. At one end of the spectrum were a number of loans which the Rothschilds launched for Brazil, where given the quality of the paper the only obligation which the bank was prepared to undertake was 'to take charge of the negotiation' of the loan.[37] On the other side of the coin, Brazil was bound to employ Rothschild exclusively for the payment of interest, dividends and the operation of sinking funds! At the other end of the spectrum a bank might agree to purchase the whole of the securities and at its discretion sell them on to investors. That only occurred where there was no risk of the issue collapsing. There are instances of this with loans, for example, to colonial governments, where the Rothschilds under the agreement contracted to 'procure purchasers'

[33] RA 000/401 A.
[34] See Exchequer Bonds Act 1854, 17 Vic, c 23.
[35] J Thabie, 'European Banks in the Middle East', in R Cameron & V Borykin (eds), *International Banking 1870–1914* (1991), p 430. cf D Platt, *Latin America and British Trade 1806–1914* (1972), p 279.
[36] A Chandler, 'The United States', in A Chandler & H Dames (eds) *Managerial Hierarchies* (1980), pp 18–19, 30–1.
[37] RA 000/401 C (1875), R 000/401 D (1886).

for the whole of the loan and pay in respect of any part not purchased.[38] Intermediate between these two positions was where the bank agreed to take a certain number of definitive bonds absolutely, but as to the remainder took only an option to be exercised within, say, a year after. These commitments on the bank's part reflected its view of the risk involved and in some cases other factors such as its moral obligations to the borrower or its anticipation about future business.

Throughout the century banks would sometimes act along with other banks in a syndicate in putting together a loan for a borrower. Indeed, taking a loan between banks was a feature of lending to the UK Government at the end of the eighteenth century.[39] Not all banks like syndicates. Leading investment banks such as the Rothschilds and the Barings were reluctant to enter them, both as a measure of prestige and because it reduced their profits. However, even they had finally accepted the inevitability of syndication by the 1860s, apparently to meet the competition when American investment banks opened in London,[40] syndicates were composed often on the basis of personal and social relationships and understandings that business acquired in certain areas was to be shared. For example, the usual partners for the Barings were Hope in Amsterdam early in the century and the Morgans in the United States later. The Rothschilds of course were able to put together a considerable syndicate based on family connection alone, the different Rothschild houses in Europe, coupled with other banks in which the family had an interest. It was clearly recognised by 1870 that syndicates enabled larger loans to be issued than otherwise would have been possible and some loans were possible which would not have gone ahead without the support of a number of banks.

In the first part of the century when syndicates were used they were put together informally without written contract or formal code.[41] However, in an article in 1876 the *Bankers' Magazine* noted that syndicates were becoming more common after 1870 and that they were being reduced to writing.

> The details of the syndicate arrangements differed in various cases, but the general plan was always the same. When the preliminary arrangements had been completed between the borrowing government, or its authorised representatives in this country, and the firm who were to undertake the introduction of the loan to the public, the latter communicated to various capitalists, bankers, merchants, stockbrokers, private individuals and others, the terms on which they could enter the syndicate. An agreement was drawn up between the original agents, or contractors, and the members of the syndicate.[42]

[38] Eg, New Zealand Loan 1875; RA 000/401 C.
[39] J Grellier, *op cit*.
[40] S Chapman, *The Rise of Merchant Banking* (1984), p 160.
[41] D Landes, 'The Old Bank and the New', in F Crouzet, W Chaloner, W Stern, *Essays in European Economic History 1789–1914*, p 119.
[42] 'A Sketch of the History of Foreign Loans' (1876) 36 *Bankers' Magazine* 517, 518.

The *Bankers' Magazine* went on to note that under the arrangements the syndicate might guarantee to take a certain proportion of the securities backing the loan, upon which their liabilities would cease, but as to what they had undertaken to place if the subscriptions fell short then each member would take a proportionate share of the balance, according to its participation. An example of a syndicate agreement is for the issue of bonds to fund an Austrian railway in 1871, promoted by the Rothschilds. The syndicate has 11 members who agree to participate to the extent of the number of bonds opposite their signature. Under the agreement all members of the syndicate are to be admitted on equal terms and conditions. However, the agreement makes clear that 'the entire management and control of the syndicate operations' are to remain in the hands of the Rothschilds. Once the bonds are distributed and paid for, and the account of profits of the syndicate made up and given effect to, the syndicate is dissolved. In the event of the purchase of bonds falling short of the numbers which the syndicate as a whole agreed to take, the agreement obliges each member to purchase in proportion to their subscriptions as to the entire number guaranteed.[43]

Despite the emergence of written agreements governing the operation of syndicates, matters were still arranged informally. An example is the agreement for an 1875 US issue of $45 million 5 per cent bonds called, incidentally, a 'funded syndicate', signed by August Belmont and Co of New York on behalf of the Rothschilds in London, Seligman and Co of New York for Seligman Bros in London and JS Morgan and Co of London (per Drexel Morgan and Co). In a side letter from Seligman Brothers to Rothschilds, however, agreeing their share and that their interest in the profits, advantages and losses were to be in the same proportion, Seligman Brothers add significantly that they 'understand that the sole management of the whole business is to remain in your [ie, Rothschild's] hands'.[44]

Another example of informality is that it might be understood by all that another bank is to be a member of the syndicate, even though its name is not on the agreement or the initial public notice of the offering.[45] Apparently security agreements often included a commitment to pay the borrower in instalments, regardless of the volume of securities sold, and also required that securities be sold at or above syndicate price. Securities were sold gradually at the syndicate price rather than being dumped on the market.[46] The cosy relationship between syndicate members, and the fact that syndicates were often overly large so that the underwriting strength was far in excess of the risk involved, gave rise to antitrust enquiries into American syndicates in the early twentieth century.[47]

[43] RA 000/401.
[44] RA 000/401 D, Letter of 27 July 1874. There is a similar letter from Morgans.
[45] V Carosso, *The Morgans. Private International Bankers 1854–1913* (1987), p 494.
[46] H Cleveland and T Huertas, *Citibank 1812–1970* (1985), p 47.
[47] V Carosso, *Investment Banking in America* (1970), pp 144–153.

III. INVESTORS

> She learned, to her horror, that Margaret, now of age, was taking her money out of the old safe investments and putting it into Foreign Things, which always smash. Her own fortune was invested in Home Rails, and most ardently did she beg her niece to imitate her.
>
> (E M Forster, *Howard's End*, 1921)

Since issues did 'smash' the problem from the point of view of legal enquiry is the extent to which the investors—be they widows, orphans, or others in the great Victorian play—might have a claim against the bank or against any security which the borrower had made available to back the loan. There may have been claims with such failed issues against the directors of issuing companies or others associated with their promotion, but these are outside the scope of the present inquiry.

A. Prospectuses

Well into the nineteenth century, the prospectuses soliciting applications for subscriptions for public loans were simple documents, often no more than one printed page. The nature of the loan would be set out, the borrower, amount and interest. Then followed the denomination of the bonds, when and where the interest coupons were payable, and any provisions about redemption. Once income tax became a feature in borrowing countries, the prospectus might note that payment in London was free of such tax. Any security would be briefly described. Since the subscription price of a bond was typically payable by instalments, over a year or so, a prospectus would include details of this together with a forfeiture provision—failure to pay all the instalments regularly would lead to forfeiture of all previous payments. It was to ensure that the forfeiture provision was effective that Nathan Rothschild had obtained legal advice early in the century.[48] Once the concept of allotment was formalised, the prospectus would make explicit that subscribers for bonds in a popular issue might not be allocated all the bonds they applied for, and provide for a return of the surplus on the amount deposited. The prospectus might inform subscribers that bonds would be exchanged for scrip receipts—issued on payment of the instalments—on or after the day the last instalment was paid.

From a legal point of view, it would have been difficult for bondholders to mount a suit on the basis of such skeletal material. With American Rails, however, the prospectuses became more detailed. Initially this was by the annexure of a letter from someone like the director of the parent company

[48] Letter from C Chatfields, Angel Court, to NM Rothschild, 10 September 1824, RA 000/401 A.

putting flesh on the bare bones.[49] For example, the 1874 prospectus for first mortgage 6 per cent sterling bonds in the New York and Canada Railway Company stated that the issue was not only authorised by, but was also guaranteed by, the Delaware and Hudson Canal Company. A letter from the canal company's president, an annexure to the prospectus, contained various representations. Some could be characterised as representations about the future, or were in the nature of forecasts, and hence not actionable in English law—for example, the line would 'open up the abundant mineral wealth of and material sources of the Champlain iron region', and when completed would cost between $6–7 million.[50] Many representations, however, were of existing fact—the nature of the corporations, and an outline of the existing assets ('the canal is first-class in every particular'; the canal company's railways, together with the New York and Canada line, aggregate over 600 miles, 'giving them the best routes for the distribution of coal and miscellaneous traffic between New York, Lower Canada, New England and the south western states of the United States'; and that the proposed security was 'first-class'). Similarly, the letter from the President of the Louisville and Nashville Railway Company, which had authorised another Barings issue the previous year—£1,100,000 sterling bonds in the South and North Alabama Railway Company—wrote that since its opening in the previous year, the latter's traffic 'has exceeded our most sanguine expectations, and will be largely increased as the large deposits of Coal and Iron upon its route become better developed'.[51]

Indeed, as well as statements by those associated with the particular development, American issues started to contain statements by purportedly neutral experts. The prospectus for 7 per cent first mortgage bonds in the Great Republic Gold and Silver Mining Company, brought to the London market in 1867 by the bankers Prescott, Grote, Cave & Co, provides an example.[52] The prospectus announces that the company is incorporated in Virginia, but has bought valuable mining land in Nevada. The prospectus has attached to it a report by a mining engineer, a letter by Professor JE Clayton, surveyor and mining engineer, and a table from the Mining and Scientific Press of San Francisco, comparing the dividends of leading companies in Nevada. Purportedly factual in part, the reports by the two experts could not have been bettered by the promoters of the company themselves. For example Murray, the mining engineer adds this puff.

> I will further add, that this district when developed will far exceed Silver Peak in said County, my reasons for making this statement will be plain to all who may visit either, namely, the geological formation is far in favour of Manhattan,

[49] D Adler, *British Investment in American Railways 1834–1898* (1970), p 55.
[50] BA, Prospectuses 29.
[51] *Ibid*, 44.
[52] NWA, d 2897.

> and the Peak, considering the capital expended, has in the space of three months given a dividend which excels the Comstock in its palmist days.

The prospectus itself invokes yet other experts. There is a reference to a Professor Sulliman of Yale College, who supposedly in several lectures on the silver mines in Nevada has expressed himself 'highly pleased with the mines in this section'. When things went wrong, however, as it did with companies such as the Great Republic Gold and Silver Mining Company, it never seems to have been thought worthwhile in legal action to join the bankers as defendants along with the company promoters directly responsible for such extravagant claims.[53]

B. Bonds

While a global bond might be signed initially by a borrower, definitive bonds always followed. These were printed to a high standard, so that despite widespread default by borrowers they become a collector's item in the twentieth century. On the face of the bond were the undertakings, along with the coupons to be detached and presented for payment of interest. The undertakings were anticipated in the prospectus. Primarily these related to the payment of interest, the redemption of the bonds and the establishment of arrangements in the nature of security. In the main, the payment and redemption undertakings were unexceptional. Although a regular feature of international lending, default was not contemplated by the undertakings, for example, by the payment on default interest. However, bonds did not always deny a possibility of catastrophe. In one Turkish issue, for example, interest was to be payable, as well as the sums constituting the principal, in times of war and peace, whether the holders be subjects of friendly or hostile states.[54]

An issue which did arise in relation to the bonds themselves was their negotiability. The issue of negotiability arose in cases as various as the bondholders suing a bank on the basis that the paper was not marketable, an issuer denying that the particular instruments were its paper, a bank having had instruments transferred to it by way of security and then trying to exercise a right of sale, and simple cases of theft and fraud. The English courts were remarkably liberal about attributing negotiability to paper. They did this on the basis not of statute but mercantile custom and usage. The great case of *Goodwin v Robarts*[55] demonstrated how flexible the common law could be, even though the type of instrument might only have

[53] Company promoters were a popular target, as in Trollope's *The Way We Live Now*.
[54] Imperial Turkish 6% loan, 1858, NWA, d 2897.
[55] *Goodwin v Robarts* (1875) LR 10 Ex 337, affd (1876) 1 App Cas 476.

come to the market relatively recently. Foreign bonds had their negotiability recognised early in the century, and later it was the turn of bonds issued by foreign companies.[56]

Despite the liberality of the English courts on the issue, the market sometimes assumed negotiability too readily. The problem arose with American railway shares which were registerable in the United States. When the shares were brought to London they were registered in the name of the arbitrage dealer or jobber who was involved in the issue. Since few American railways maintained registry offices in London, the share certificates would have to be returned to the registry office in the United States when they were bought and sold, but few purchasers troubled to do this. The Stock Exchange evolved a method by which these registered shares passed by transfer just like bearer bonds. The practice grew up of the registered owner signing the blank transfer form on the back of the shares, but leaving the name of the transferee blank. The market assumed that the instruments were negotiable.[57] However, when the issue was finally brought before the English courts in the late 1880s, it was held that the certificates were not intended to be passed by mere delivery and were not negotiable instruments.[58]

Bonds were more popular with investors than shares. Share frauds in the early 1850s, in American stocks, a failure by companies to declare regular dividends, and regular calls on shares led investors to prefer the fixed interest of bonds, especially since they, unlike shares, might be backed by security.[59] However shares were issued, including bearer shares, and in jurisdictions as various as Mexico, Russia and of course the United States. Convertible bonds were also started in the nineteenth century, whereby at the option of the investor the bonds could be converted into fully paid up shares of the company. They provided an added attraction to investors, who if the venture was a success would be able to take advantage of its growth. The method for convertible shares was to transfer shares from abroad, and place them at the disposal of trustees in London, who would convey them to investors once the bonds were submitted for conversion.

Investors were disadvantaged by market rigging, stagging and the general fraud that occurred on the London Stock Exchange in the nineteenth century and gave rise to Parliamentary enquiries and a Royal Commission.[60] These practices are outside the scope of the present enquiry. However, banks did engage in support activities either on the issue of paper or in the

[56] Government bonds: *Gorgier v Mieville* (1824) 107 ER 651; *A-G v Bouwens* (1838) 150 ER 1390; *Lang v Smyth* (1831) 131 ER 109. Corporate Bonds: *London Joint Stock Bank v Simmons* [1892] AC 201; cf. *Glyn v Baker* (1811) 104 ER 468 (East India bond).
[57] D Adler, *op cit*, 177.
[58] *London & County Banking Co v London & River Plate Bank* (1887) 20 QBD 232; *Colonial Bank v Cady & Williams* (1890) 15 App Cas 267.
[59] D Greenberg, *Financiers and Railroads 1869–1889* (1980), p 24.
[60] Eg, D Kynston, *The City of London: A World on Its Own, 1815–90* (1994), v 1.

secondary market, in order to support the price.[61] Making a market, or stabilisation as it is now called, was probably the most crucial activity for a bank involved in an issue, since it was necessary to create enough interest or appearance of interest in the paper to induce investors, but also to convince the London Stock Exchange Committee that a respectable proportion of the issue had been allotted so that there could be a listing if this was desired. Possibly influenced by the adverse comments of the Royal Commission into the Stock Exchange, the Court of Appeal when it first considered stabilisation held that it rendered a contract unenforceable through illegality.[62] Shortly after, however, in another decision, the Court took a more favourable view and held that stabilisation was unobjectionable if honestly done, within limits, and in accordance with market practice—and, most importantly, the fact that stabilisation was to be done was publicly disclosed.[63]

C. Security

We have seen that state borrowers might undertake for a particular period not to increase the public debt without the consent of the banks. Even when incorporated in legislation, that undertaking was not enforceable either by the banks or investors. State borrowers might also undertake that the sums payable under a loan were payable from or specially charged on the proceeds of, for example, the rent from lands of the state or from particular revenues, for example customs at a particular port. Again it is difficult to see how this was enforceable. The matter arose in relation to a Peruvian loan where the Government had said—in a statement on the bond—that 'as a guarantee for the fulfilment of the obligations contracted in this bond, the Government of Peru, under the national faith, pledges that general revenue of the Republic, and especially the free proceeds of guano in Europe and America after the engagements which it has contracted on them are covered', and that 'in all contracts which the Government may enter into for the sale of guano, or under whatever form the sale may have, it binds itself to direct that there be set aside out of the proceeds' money for the payment of the bonds. The English Court of Appeal, in an action by a bondholder against the agents of Peru in England, held that the bonds were engagements of honour only, and did not contain any contract enforceable by an English court. Moreover, and perhaps surprisingly, the Court of Appeal held that the bonds were not enforceable by the courts of Peru as well, without the consent of the Government.[64] When a foreign government's

[61] C Goodhart, *The Business of Banking, 1891–1914* (1972), pp 120, 124. The practice was introduced from England into the United States by Jay Cooke in the 1860s.
[62] *Scott v Brown, Doering, McNab & Co.* [1892] 2 QB 724 (CA).
[63] *Sanderson and Levi v British Westralian Mines and Share Corp Ltd,* The Times 19 July 1899.
[64] *Twycross v Dreyfus* (1877) 5 Ch D 605.

issue was guaranteed by the United Kingdom Government, however, as happened in the case of an Egyptian issue in 1885, then the investors had a stronger claim.[65]

Those taking paper issued not by states but by companies had a stronger legal claim, even if in practice this did not produce results. First, there might be a guarantee. The guarantee of the Russian Government of an issue by a Russian railway company in 1857 would have faced the problems of enforceability already mentioned in relation to government issues.[66] However, the guarantees given by US parent companies with railway issues would have provided greater assurance. Secondly, corporate issuers might give security, enforceable by trustees of the issue. American railways, which were built with land grants from the state, had an especially valuable form of property which could be secured, and from the middle of the nineteenth century banks started to act as trustees of American railway bond issues. The security for issues covered not only the land but also other physical assets such as moveables like the railway cars and locomotives.

Very early in the nineteenth century Anglo-American law displayed a receptivity to security not evident in continental European countries. This was especially the case in the way security was permissible over intangibles such as bonds and shares. In an Austrian loan of the 1790s, Bank of Vienna debentures had been deposited as a form of collateral security.[67] Bankers from the start of the nineteenth century regularly advanced moneys on the basis of shares and bonds deposited with them, both in England and the United States. Where the securities were bearer instruments it was easy for common law courts to treat this as a pledge. To do this with the deposit of non-bearer share certificates was more difficult conceptually, although there were a few cases where the pledge analogy was used.[68]

IV. DEFAULT

> An Eagle stayed his flight and entreated a Lion to make an alliance with them to their mutual advantage. The Lion replied: 'I have no objection, but you must excuse me for requiring you to find surety for your good faith, for how can I trust anyone as a friend who is able to fly away from his bargain whenever he pleases'.
>
> (Aesop, used as a subscript for a cartoon by Thomas Nast in August 1876).

[65] Egyptian Loan Act 1885.
[66] Grand Russian Railway Company, Issue of 100,000 shares in London, 28 April 1857, BA. The guarantee was in an Imperial Ukase (decree) of 26 January 1857 (old style) attached to the prospectus.
[67] S Cope, *Walter Boyd: A Merchant Bank in the Age of Napoleon* (1983), 51, 53.
[68] These earlier cases were distinguished in *Harrold v Plenty* [1901] 2 Ch 314.

Default by borrowers was not of direct concern to bankers in the first part of the nineteenth century. The loss fell on the bondholders. While the banks themselves may have taken part of the issue in the initial stages, they tended not to keep the paper, certainly once the borrower got into difficulties. However, bondholders were sometimes able to exercise an informal power to have the issuing banks take action against a borrower. Thus with American Rails, banks in London put pressure on their affiliates in the United States to act to protect the interests of their bondholding clients. To an extent American bankers were willing to assist because the continued flow of foreign capital into American Rails might otherwise be jeopardised. The bondholders would appoint the banks as their agents to effect a settlement. The banks then used this to put pressure on railway executives. In some cases American bankers actually became involved in running railways and in extreme cases effected a restructuring by the sale of the railway or its merger with another line.[69]

Restructuring was facilitated where mortgage bonds were involved. The bondholders were then in a position to exercise their security to seek a sale. Where this occurred the courts were supportive. Thus in 1876 a US Federal Court ordered the foreclosure and sale of a line. This result was so encouraging that foreign investors were advised to turn to the American courts whenever there was anything to be salvaged.[70] At the foreclosure sales, it was possible for bondholders to purchase the property themselves. In some cases this was done and the bondholders ran the railway through an American bank; in other cases the sale was used to purchase the line and then lease it to another operator. Defaulting railways, and their restructuring, saw the development of new doctrinal law in the United States ('equity receivership') and of an elite group of lawyers to effect it.[71]

It was surprising however that where a reasonably satisfactory outcome like a restructuring was impossible action against the banks themselves does not seem to have occurred. After all, the banks promoted some very shoddy paper, and were handsomely rewarded. One reason is that early in the nineteenth century the banks saw the need to insulate themselves from claims by bondholders. Sometimes clauses were inserted in agreements between bank and borrower which exempted the bank itself from liability. While the bonds represented claims against the issuer, the banks would have been able to use this exemption where a borrower defaulted.

Secondly, it may have been that where disputes arose they were dealt with informally and did not attract attention. Thus when the Duc d'Otrante

[69] Eg, R Swaine, *The Cravath Firm*, v 1, contains many examples.
[70] D Greenberg, *Financiers and Railroads, 1869–1889*, p 121, 150, J Storer, *The Railroads of the South 1865–1900* (1955).
[71] See G Glenn, 'The Basis of the Federal Receivership' (1925) 25 *Col Law Review* 434; D Skeel, *Debt's Dominion. A History of Bankruptcy Law in America* (2001), ch 2; D Baird & R Rasmussen, 'Control Rights, Priority Rights and the Conceptual Foundations of Corporate Reorganizations' (2001) 87 *Vir Law Review* 921.

complained about a reduced dividend on American railway to Barings Lord Revelstroke wrote: 'I do not wish to avoid any responsibility I may have incurred in recommending the investment to you and you will today receive a letter from my firm crediting you with the sum you paid for the bonds. As however you think proper to say you think the whole thing is a 'swindle' I should prefer that our business relation should cease and I have to request that you will as soon as possible inform my firm to whom they can transfer the stocks and money they hold for your account.'[72]

The converse of borrower default—the failure of the bank—sometimes gave rise to problems. In 1837 Thomas Wilson & Co ran into difficulties and suspended payment. It had arranged a loan for the Danish Government in 1825 and it continued as agent for payment of interest (on the presentation of the coupons) and for the sinking fund—a valuable appointment for which commission was payable. Mindful that the moneys if remitted half yearly to London might disappear were the bank to become insolvent, Denmark wished to terminate the agency and to appoint Rothschilds as the new agent. Thomas Wilson & Co, hopeful that they might pull through, stood on their rights, since there was nothing about termination of their agency in the loan agreement and their appointment as agent was part of the consideration of the original loan agreement. Denmark suggested to Rothschilds what would be expected in the modern age in this type of difficult situation—that they obtain counsel's opinion, no doubt in the hope that this would be favourable and intimidate the other side. Who better than Sir Frederick Pollock, who had shortly before been Attorney General, who was to go on to the Bench as Chief Baron, and whose sons and grandsons made notable contributions as judges and academics to English law? Pollock's opinion of 9 August 1837 covers a little over a page in manuscript. Without citation of authority it invokes that happy refuge of the lawyer when faced with an unhelpful written contract, the implied term.

> It must be an implied condition with regard to every duty or office that the party claiming to discharge it should be fit and competent to do so. Messrs Wilson & Co are not now fit and competent and I think therefore they cannot claim the performance nor is the Danish Government bound to employ them or accept as a substitute any other agents named by Wilson & Co or any security for the correct appropriation of the funds to be remitted.[73]

As to Thomas Wilson & Co's point about consideration, Pollock quickly disposed of this by drawing the distinction between motives for entering, and the legal consideration underlying, a contract. In the result the agency

[72] P Ziegler, *op cit*, p 218.
[73] RA 000/401.

was transferred to Rothschilds. Thomas Wilson & Co became insolvent shortly after.

Once default had occurred, it was impossible for a period for the state or company to raise new money. This reputational effect sometimes spread more widely to a region. American states were in very bad odour for a number of years after defaults in the 1830s. Eventually, however, new loans were possible with a change in government policy, the advent of a new government or in the emergence of an especially attractive project to be funded. Repayment of the loans in default might be necessary to attract new money. In 1858 Barings announced to holders of Buenos Aires bonds that it would exchange for new bonds the unpaid interest coupons due from 1830 and never paid. Fresh coupons for the old bonds would also be issued. The Committee of Spanish American Bondholders had been involved in the rescheduling and had convened a meeting of old bondholders two years previously. The Barings announcement contained an extra resolution of that meeting, authorising a deduction of 5 per cent from the amount of arrears to go to the Committee for their remuneration. Barings, the announcement added, would not receive any part of this and would not be charging for its services for the issue of the new bonds. Sometimes the need to maintain reputation extended to the bank as well! Under a rescheduling agreement entered by the Governor, Buenos Aires undertook to remit specific amounts to Barings to pay interest on the bonds and for a sinking fund for their repurchase.

V. CONCLUSION

Globalisation as an analytical concept may be new, but the phenomenon it describes is not. The period 1800–1914 was one of intense globalisation. Partly this was because this was an era of colonisation and empire, but only partly. Migration of people and cultures, cross-border trade and, in terms of the present paper, capital flows, all contributed. It was not for another half century that the massive capital flows of the period ending in 1914 were in any way matched. That required the establishment of the IMF and World Bank as part of the Bretton Woods settlement, enormous foreign direct investment in the 1960s and the development of the Euromarkets from the early 1970s.

The major argument of this chapter is that the legal techniques and institutions of modern international lending have their grounding in nineteenth century. Bond issues, syndication, and the remedies on default are the main examples. This is not to say that everything was set in stone with nineteenth century international lending or that nothing new occurred subsequently. Domestic practice in the period 1918–1970, especially in the United States, was subsequently moulded to the needs of the Euromarkets. For example, syndication as an agreement between bankers had a long history. However, syndication is also possible by 'sale' of the loan, and that developed first in American domestic practice.[74] But a discussion of American legal techniques, and their extrapolation to international lending, is for another day.

[74] Eg, 'Participating Mortgages as Method of Trust Investments by Corporate Fiduciaries' (1936) 45 *Yale LJ* 857.

Commentary on 'Globalisation: Its Historical Context'

CATHERINE NEWMAN, QC

ROSS CRANSTON'S CONCEPT is that globalisation, whilst a relatively new word, is a trend which has been developing for the last two centuries. Most modern lending techniques have their roots firmly in the nineteeth century. Ross Cranston reminds us that before World War I the flow of money was chiefly from Britain, France and Germany to the rest of the world. That dominance was extinguished by the debt burdens incurred during the Great War and by the continued growth of the role of the United States of America in world lending.

Globalisation is not now a term of approval: it is today's target for protest. For many it means 'Americanisation' and the fabulous wealth and power of America's big corporations. The globalisation of terrorism and the global effects of pollutants on the world are beyond the scope of this book. We are looking at one highly significant aspect of globalisation: the globalisation and development of money-raising techniques.

The 1999 Reith lectures delivered by Professor Anthony Giddens of the London School of Economics focused on issues arising from global capitalism. He used the term 'runaway world' to describe an environment which we cannot control. Certainly control or regulation and the availability of satisfactory remedies at different levels is a complex problem which must be grappled with. When global lending and restructuring fails, we have global insolvency.

Historically in the United Kingdom the emphasis in such a situation has been on the protection of creditors' rights, though with the Enterprise Act there is a shift in direction. In France insolvency proceedings try to focus on saving the business and the maintenance of employment. As in Germany, in France the court will appoint administrators who are supposed to run the business. Administration Orders in England and Wales and, in the US, chapter 11, run in ways which are in some respects similar. Japan has a process of composition and reorganisation. South American countries have a variety of procedures. In a global collapse secured creditors wanting to foreclose may find themselves in a rather different position depending on which national card is the first to fall out of the pack.

Sovereign debt restructuring lacks the ultimate sanction of liquidation and does not, generally, offer the opportunity to replace the management team. Renegotiation of terms is not always legally possible or practical. Exchange offers can help and have worked in small countries with fairly simple debt structures. The tendency to involve major banking houses, so familiar in the nineteenth century, continues today.

Collective action clauses in bonds allowing for restructuring on a majority vote are now being used in the London market but of course are also present in many older issues. These give huge power to influential cliques of bondholders and there may be a case to be made for an international statutory regime; but that is another argument.

Ross Cranston queries why, in the nineteenth century, the banks issuing paper which turned out to be shoddy were not themselves sued more regularly. This may be because the banks issued the paper as agents only. Government agents, in those days, did not incur personal liability. They could not therefore be sued. There was no distinction made between agents for the British government and foreign governments. Agents for private persons could be sued. The bank issuing shoddy paper for a private mining or railroad company did not enjoy like immunity.

In *Goodwin v Robarts*,[1] a decision subsequently confirmed on appeal, scrip issued by Rothschilds as agents for the Russian government and the Austro-Hungarian government was bought by the plaintiff through a stockbroker and left with the broker. No criticism was made of the paper itself, but the broker unlawfully pledged the stock with the defendant bankers as security for a loan to the broker himself. He became bankrupt and did a runner. The banker sold the scrip in the market. Could the plaintiff, the original buyer of the scrip, recover the proceeds from the banker? The scrip was for £100 of stock, £20 paid and the balance due on various dates. On payment in full the bearer of the scrip was entitled to a bond. The whole had been paid up. The five-man Court of Exchequer Chamber led by Chief Justice Cockburn heard the matter argued and rejected the notion that Rothschilds could be treated as principals. Cockburn CJ said:

> … the natural presumption in such cases is that the contract was made upon the credit and responsibility of the Government itself, as possessing an entire ability to fulfil all its just contracts, far beyond that of any private man, and that it is ready to fulfil them not only with good faith but with punctilious promptitude, and in a spirit of liberal courtesy. Great public inconvenience would result from a different doctrine, considering the various public

[1] (1875) LR 10 Ex 337.

> functionaries which the government must employ in order to transact its ordinary business and operations; and many persons would be deterred from accepting of many offices of trust under the government, if they were held personally liable upon all their official contracts.[2]

And later:

> If an agent, on behalf of government, makes a contract and describes himself as such, he is not personally bound, even though the terms of the contract be such as might, in a case of a private nature, involve him in a personal obligation .... But the agent in behalf of the public may still bind himself by an express engagement, and the distinction terminates in a question of evidence.[3]

They found that the bank had not become personally bound and the scrip was negotiable in accordance with market practice. So the plaintiff lost his money and the court was distinctly unsympathetic:

> ... it is obvious that no injustice is done to one who has been fraudulently dispossessed of scrip through his own misplaced confidence, in holding that the property in it has passed to a bona fide holder for value, seeing that he himself must have known that it purported on the face of it to be available to bearer, and must be presumed to have been aware of the usage prevalent with respect to it in the market in which he purchased it.[4]

The flow of investment abroad is reflected, in legal history, by the mining and railway scandals which gave rise to so many cases of significance. Ross Cranston's paper refers to some of these and as we read we feel certain that soon we will come upon the facts of the South Central Pacific and Mexican railway scandal: the great railroad was to run from Salt Lake City to Vera Cruz; the shares were to be issued at San Francisco and in London. But the railway was never built, and the investors lost their money. With great enterprise comes the risk of fraud and failure. The Limited Liability Act of 1855 was the nineteenth century's Enterprise Act: designed to reduce the risk of failure by permitting trade with limited liability it permitted the growth of the Stock Exchange, but also encouraged Baron Grant (born plain Albert Gottheimer) to make his lists of widows and clergymen who might succumb to a prospectus offering a great return. Risk may dull enterprise, but enterprise without risk creates a climate in which fraudsters flourish. Global business needs a system of ethics and regulation capable of standing up to global fraud.

[2] *Ibid*, 344.
[3] *Ibid*, 345.
[4] *Ibid*, 353.

2

Commercial Notions and Equitable Potions

SIR JOHN MUMMERY

SOME WISH THAT they could have attended the late Professor Nozick's philosophy class at Harvard on 'The Best Things in Life.' The course was an exploration of the nature and value of those things deemed best in life: friendship, love, intellectual understanding, sexual pleasure, achievement, adventure, play, luxury, fame, power, enlightenment and ice cream.

Imagine a course on the nature and value of those things deemed best in English Law. There would be no shortage of ideas or materials. A millennium of English Law has produced a 'treasure house.' Over the centuries amazing stories have been told in court. The stories have been re-told by judges. They have tried to make legal sense of them. Their judgments have been studied and analysed by legal writers. In their search for meaning and coherence in the law, they increase our knowledge and improve our understanding.

I. THE TRUST IDEA: INTRODUCTION

This is a superficial treatment of scattered topics in a complex subject. It suggests that the best things in equity can be used for a more ethical framework of conduct in some of the areas covered by commercial law. The discussion concentrates on the 'trust idea'. FW Maitland[1] described its development as English Law's 'greatest and most distinctive achievement.'

The 'trust idea' is strong. It has endured, expanded and adapted over many centuries. It is a simple and practical legal concept reflecting the realities of life and the moral and social values generally accepted in society. Trust of one another, relying on others to act in certain ways expected of them, operates on many levels in public and private life. Trustworthiness cements

[1] *Maitland: Selected Essays*, H D Hazeltine, G Lapsley and PH Winfield (ed) (Cambridge University Press, 1936) at p 129.

personal, social and economic relationships and institutions. It is a necessary condition of all cooperative activity. It is fundamental to the workings of human society. It is taken for granted, so much so that its value is not appreciated until it is too late. Almost without anyone bothering to lift a finger, it can slip away. Its disappearance does untold damage in social, human and economic terms. Everyone knows from personal experience that, once lost, trust is not easily regained.

The 'trust idea' is larger then the 'trust institution'. The trust, as a juristic institution, has dominated equity for the last 300 years or so. For as long as land was the principal source of personal wealth in England, much of it, which was not the property of the Crown or the Church, was held in trust. The trust directly enabled many landed families to preserve the family fortune for centuries in strict settlements and indirectly formed the social and political order from the late seventeenth century until the early twentieth century.[2] As the forms of wealth have changed and been re-distributed, so have the uses of the trust.

As Professor PS Atiyah has written, however, the fact of trust preceded the institution of the trust.[3] The institution of the trust owes its origins and development to the ethical response of equity to the fact of broken trust. Equity regards it as wrong for a person to use for his own benefit property entrusted to him for the benefit of other persons or for specified purposes. As a result of equity's recognition of the social phenomenon of trust and of the need to enforce it, the trust idea evolved into the trust institution. The setting up of legally enforceable trusts was made possible.

In English Law the trust idea preceded bargains as a source of legal obligation. In fact, bargains usually operate on the basis of mutual trust and are not dependent on the legal enforcement of contractual obligations. It fortunately still is possible to achieve a significant degree of effective cooperation in commercial activities on the unspoken assumptions of mutual trust, without the need for legally enforceable promises and without recourse to litigation or the threat of it.

II. COMMERCIAL LAW AND THE TRUST IDEA

Although promises and contracts will always remain the primary concepts of commercial law, more extensive and imaginative development of the trust idea should be made (a) to improve standards of conduct in commercial relationships and transactions and (b) to achieve more just outcomes in the resolution of commercial disputes.

[2] Sir John Habakkuk, *Marriage, Debt and the Estates System: English Land Ownership 1650–1950* (Oxford: Clarendon Press, 1994).
[3] P S Atiyah, *Promises, Morals and Law* (Oxford: Clarendon, 1981) at p 119.

Provided that the trust idea is used sensibly, it can supplement the role of contract law in improving commercial integrity and confidence. Integrity and confidence form the essential foundations of stable and successful markets. As Amartya Sen,[4] winner of the Nobel Prize in 1998 for Economic Science, has explained, the success of capitalism is not dependent solely on competition and cupidity. As well as the self-interested pursuit of maximum profit, there is a need for self-interested altruism. Reliability, business honesty, honour, a sense of obligation and implied mutual trust are all necessary values in a market economy. Society generally, and exchange economies in particular, operate on a basic presumption of mutual trust. In commercial, as well as social, dealings with one another people are reasonably entitled to expect some guarantees of openness, disclosure, lucidity and transparency to deter corruption, underhand dealings and financial irresponsibility. Institutional structures and codes of conduct are based on the need to sustain mutual trust and confidence in each other's business ethics.

In short, I think that there is much to be said for the idea that in commerce, as well as in the rest of life, it is in your own ultimate interest to put your own immediate interest in second place. That is not far removed from the trust idea (or from some other great ideas about how we should live our lives).

The trust idea can be used in commerce without adversely affecting the ability of businessmen to create personal rights and duties by contract and without diminishing legal and actual respect for their bargains. There are, however, obvious limits to the contribution that equity can make to commercial law. Use of the trust idea must (a) recognise the primacy of contract law; (b) make allowances for unavoidable risks inherent in commercial life; and (c) adjust appropriately to the predominant role of commercial law in the establishment, operation and expansion of stable markets and in lending certainty and predictability to commercial transactions.

III. DEVELOPMENT OF THE TRUST IDEA IN COMMERCIAL SETTING

Legal practitioners and the courts now have a better understanding of the considerable scope, in both contractual and non-contractual commercial settings, for making more effective use of the trust idea.

Use of trust principles, structures and techniques is now widespread in commerce. As explained in a paper produced by the Financial Law Panel:[5]

> Trust principles have long been recognised as useful in commercial relationships. In particular, the ease with which the legal ownership of assets may be

[4] AK Sen, *Development as Freedom* (Oxford: Oxford University Press, 1999) at pp 39–40, 263–267 and 279.

[5] Financial Law Panel, *Financial Dealings with Trustees: A Preliminary Discussion Paper* (October 1997, London).

> segregated from the underlying beneficial interests and the ability of different groups to have concurrent but different interests in the same property underpins many of the flexible structures created under English law. There is a large number of different ways in which trust structures are used in commercial situations and this number increases constantly.

Examples are pension funds, the most common form of investment vehicle to take the form of a trust; many other forms of collective investment scheme; commercial security devices in the form of secured syndicated lending, project financing and the issue of secured loan stock; and share ownership schemes. Trusts are also commonly employed in commerce as a means of conferring benefits on non-contracting parties, as in placing and underwriting agreements and in mergers and acquisitions.

Although the traditional areas of application of the trust idea are very different from those of commercial law, it is now generally recognised that there is scope in commercial law for using the trust idea to establish a wider range of fiduciary relationships and fiduciary duties, in taking prudent measures to deal with unacceptable conflicts of interest in financial markets, in developing a framework of higher standards of fair and honest dealing in commerce, and in protecting valuable information communicated and acquired in commercial confidence.

There are several reasons for the current receptive attitude to equity. In a lecture to the Chancery Bar Association, Lord Millett referred to the

> … growing complexity and professionalisation of commercial life which have accompanied the change from an industrial to a service economy and the growth of the financial services industry …[6]

The need for relationships of trust and confidence in commerce has grown. Recent events in the financial markets discussed later demonstrate the importance of encouraging, even imposing, higher standards of conduct in commerce.

Legal advisers are themselves more aware of the battery of personal and proprietary remedies available for breach of equitable duties: setting transactions aside, restitutionary proprietary and personal remedies and equitable compensation. This feature of equity has multiplied and popularised claims in commercial disputes based on breaches of equitable obligations.

A blurring of traditional boundaries between commercial and professional activities, between customers and clients, is taking place. The professions have taken on some of the characteristics of commerce. The supply of professional accountancy services and legal services is more commercialised. Some areas of the financial services sector now supply professional type services and engage

[6] 'Equity's Place in the Law of Commerce' (1998) 114 *LQR* 214 at p 216.

in activities very different from ordinary commercial arm's length dealings: receiving information in confidence, supplying advice and other services, acting for clients by guiding them in the management of their finances and affairs. To a significant extent, though usually short of positive actionable representations and binding promises of trustworthiness, the trust idea features in the methods used by commercial men to win clients and in handling their affairs.

Although there is a tendency in commercial 'spin' to mimic the vocabulary of trust and confidence ('trust', 'equitable', 'fidelity' and so on), there are failures to act in the spirit of the trust idea: for example, in not making proper disclosures of possible conflicts of interest arising in the services offered. If made, the disclosures might have affected the decision of the client to deal with them. It is unusual for express and enforceable promises of trustworthiness to be made. It is unusual, in the absence of express contractual terms, for liability to be admitted and accepted for falling short of standards of trustworthiness, reliability and fair dealing, which the clients may have been reasonably led to expect and to rely on.

Overall my perception is that some parts of commerce have moved into areas traditionally occupied by the trust idea, rather than the trust idea intruding in areas of commerce, in which it has no proper place.

IV. DECLINING TRUST IN FINANCIAL MARKETS

In 2002 a lot was said and written about trust generally, and mainly about the decline of it in the financial services sector. Consider a few recent examples.

In the Business Section of *The Times* extracts appeared from *'Take on the Street'* by the former Chairman of the Securities and Exchange Commission, Mr Arthur Levitt.[7] His subject was Wall Street and its dealers. The headline of the first extract was 'Trust your instinct, not your broker.'

He described the commercial role of brokers: sales people, who are paid a commission on transactions in accounts managed by them. He identified 'broker conflicts', quoting from Mr Warren Buffett, 'one of the smartest investors,'

> A broker is not your friend. He's more like a doctor who charges patients on how often they change medicines. And he gets paid more for the stuff the house is promoting than the stuff that will make you better.

If only levels of early reward determine behaviour, trust goes into serious long-term decline. As described by Mr Levitt, the system, which inspired his maxim 'trust your instinct rather than your broker', is a conflicted culture: some full service brokers are tempted to operate on the basis of selling

[7] 25 September 2002.

anything to anyone under the pressure of higher rewards from volume, rather than by what is appropriate for the investor, who lacks independent objective advice and is unaware of the commissions and bonuses involved. Although on-line trading on the internet is popular, as it appears to be easy and saves commissions, it is, as he explained, little better, as there are hidden trading costs.

Mr Levitt advised dispensing with a broker altogether. Instead, have an independent investment adviser, whom you would have to pay, but who would be under a fiduciary duty, that is, a duty to put your interests before his own and a duty to make full disclosure of conflicts of interest, which might affect the independence of his advice and recommendations.

The headline of the second published extract was 'Calamitous Loss of Trust in Wall Street.'

Mr Levitt criticised alleged conflicts of interest in large full service brokerage firms. They stand accused of recommending to investors shares in companies in an investment banking relationship with the firm. Brokers privately derided (in very colourful language) the very shares that they were recommending investors to buy. This practice spells disaster for investors, but it wins business for the brokers from the investment bankers, who undertake company flotations, mergers and acquisitions. Investment analysts employed by the brokers are alleged to act as an adjunct to the investment banking business, not as a source of unbiased decisions or advice for private investors. In some cases it is alleged that the analysts downgraded the stock of those companies, who had gone elsewhere for their investment banking services. The non-disclosure of these conflicts, the lack of transparency and openness, the tainting of research and the clouding of judgement by bias inherent in the conflicts of interest, all encourage a 'web of dysfunctional relationships between analysts and corporate clients of the firm.' This is disastrous for the outside investors. At the same time as they suffered losses, some private outside investors saw insiders apparently avoiding losses and even profiting from privileged information.

So far as the investors are concerned they have lost share value in the short term. In the long term they have lost trust and confidence in, and respect for, the financial services industry, in which, if it is to survive and prosper, trust and confidence are essential. The real long-term losers are likely to be the firms, who have reaped short-term benefits from abuses of the trust placed in them. The consequent damage to reputation and goodwill is incalculable. Trust cannot be restored by a blitz public relations exercise.

Wall Street is not alone in attracting criticisms of loss of trust and of unacceptable and undisclosed conflicts of interest. In a scintillating essay, '*The Deficit Millionaires*',[8] Julian Barnes described the breaking down of trust in the Lloyds affair and in the Maxwell pensions affair. He described

[8] J Barnes, *Letters from London* (London: Picador, 1995).

the Lloyds investors who felt betrayed. They were convinced that they had been sold the concept of an honourable society operating on trust. In return they had given their trust, as well as their money. In the memorable words of one commentator, they had put 'rabbits in charge of the lettuce.'

Problems of conflict management have also surfaced in the wake of the Enron affair. It has put under the spotlight the conflicted role of accountants acting as consultants to audit clients. One leading investment banker said of conflicts of interest within accountancy firms:

> I have a simple statement: I do not believe that auditors should provide consulting services to their audit clients because, at least in appearance, this compromises auditor independence and erodes trust.

The banker may not have welcomed a journalist's riposte:

> I have a simple statement: I do not believe that investment banks should provide research services to both their investment and corporate clients simultaneously because, at least in appearance, this compromises research independence and erodes trust.

On a wider front another newspaper headline in 2002 read '*Trust me, I'm giving a Reith lecture.*' This was a reference to Lady O'Neill's Reith Lectures, '*A Question of Trust*.'[9] They were a timely examination of the lessening of public trust and confidence in institutions generally. She described the universal imperative of trust—for each individual, for every institution and profession and for society—and of the crisis of trust in public institutions, in big business and its products, in banks and financial institutions, in pension providers and so on.

In the discussion following the first lecture a story was told of a woman who had made a great deal of money by acting on a financial tip in an internet chatroom. When the source was tracked down, it was discovered that a 14-year-old schoolboy had posed as a financial adviser. Two other schoolboys were also dispensing advice, one posing as a priest advising on moral questions and the other posing as a surgeon advising on medical matters. They were all impersonating people who would, in normal circumstances, be regarded as professionals, and in whose trustworthiness you could place complete confidence.

So what happens when internet technology can devalue professional trust and decentralise institutional validity? The questions then tend to become more and more extreme. Can you trust technology? Can you trust people who pose as trustworthy? Who can you trust? Can you trust anyone?

[9] O O'Neill, *A Question of Trust* (Cambridge: Cambridge University Press, 2002).

V. RESTORING TRUST IN MARKETS

The questions and the answers should not induce a state of panic and negativity. They should lead to a constructive and positive question: how is it possible for trust in the market, once broken, to be restored, maintained and promoted?

Richard Lambert wrote in *The Times*[10] an article on the betrayal of the trust of investors and the loss of confidence of investors in the integrity of the markets and those operating them that 'trust, once betrayed, takes a long time to be rebuilt.'

For trust to be restored those in whom trust has been lost must act and, just as importantly, appear to the world to act, in a trustworthy way in their dealings with others. This may require a major re-organisation of institutions in order to reduce unacceptable conflicts of interest, which have undoubtedly contributed to the problem.

As another financial journalist wrote in *The Guardian*:[11]

> Well, the business and financial communities have got themselves into the mess, in which ordinary investors no longer trust the system and the consequential buyers' strike is weighing down the system. Self-help is probably also the only credible way out. It's just going to take a very, very long time. There is no simple panacea.

But self-help is only a partial solution. Whether done internally or externally the apportioning of blame and the re-allocation of responsibilities do not by themselves solve the problem of a fundamental breakdown in trust and confidence. This is not simply a matter of finding a way of compensating for losses or of passing on the costs, which have been suffered from the failure of trust. Nor is it a mere matter of litigation, obtaining judgments against wrongdoers or putting criminals in the dock. Far more fundamental changes must be made in individual attitudes, as well as in laws and organisations. No progress will be made without a general acceptance of individual responsibility, of being personally accountable and answerable for living up to reasonable expectations of trustworthiness.

The truth is that we are all affected in one way or another by the decline of trust in institutions. It is not simply someone else's problem for them to solve. Collaboration is called for. We must all address the questions: what sort of changes are possible to restore trust? And how best are they to be achieved?

If appropriate steps are not taken voluntarily, changes may have to be imposed by legislation and supervised by regulatory authorities. There may have to be organisational splits, voluntary or enforced, separating,

[10] 8 October 2002.
[11] 8 June 2002.

for example, investment research from other aspects of banking and brokering, separating auditing services from the supply of other services.

Actions speak louder than words. Words are also required. Contract comes in at this point. If people in commerce wish to be trusted by others, they must be ready and willing to make binding and meaningful promises to act in a transparent and trustworthy way, to engage in client candour and concern, as well as to act with competent care.

But even contracts may not be enough. They are not always necessary or sufficient for the courts to enforce appropriate standards of trustworthy conduct. Equity can make a modest contribution at this point by supplying a framework for the observance and enforcement of higher standards. Equitable obligations are not dependent on a legally binding contract between the parties. The obligations can arise, for example, from circumstances giving rise to fiduciary relationships and generating the duties of a trustworthy friend, as well as those of a careful and competent neighbour.

These obligations stem from the core notion that equity acts on the conscience of the individual, as in the paradigm case of the holder of the legal interest in property, and imposes obligations to curb and cure unconscionable conduct.

VI. COMMERCE AND CONTRACT

As explained earlier, in a general, non-legal sense trust is essential to commerce. A leading merchant banker said, 'Integrity is the cornerstone, if not the bedrock, on which all financial markets are based.' This should be true of all markets, whether in commodities, securities or financial services. Trust is placed in the integrity of the market and in those operating within it. If it is misplaced, trust will go.

Promises are usually made and exchanged in a spirit of trust, in the expectation that they will be kept rather than broken. Trust is, in that general sense, put in the efficacy of the law, of contract in particular, and, as a last resort, in the courts performing their function to administer justice. Although the courts cannot make a contract for the parties, they will attempt, wherever legally possible, to give effect to an ascertained and settled intention on the part of the parties to a contract, by upholding concluded commercial bargains and by resolving difficulties, uncertainties and ambiguities in their dealings and communications by construing them fairly and broadly, so as to accord with their probable intentions.

Sometimes a commercial contract will contain express provision for the creation of a trust of identified property for ascertained beneficiaries, who may or may not be parties to the contract, or for specific purposes. In some cases it may be possible to construe the contract concluded between the parties as evincing a clear intention to create a trust of property, including a trust of a promise, even though the word 'trust' has not been used.

There is, of course, no reason why the parties to a commercial contract should not, short of creating a trust, agree express terms under which a party, in the performance of the contract, is bound by duties of a kind normally binding on express trustees. In appropriate cases, depending on the type of contract, the relationship between the parties, their intentions and all the relevant circumstances, the court implies terms into a contract for trustee type obligations of loyalty, disclosure and confidence in order to give efficacy to the contract as, for example, in the case of a contract of agency. This would not be easy in the case of an ordinary commercial or business transaction. In applying the test whether the proposed implied term was necessary to give business efficacy to the contract, the court is bound to consider what must have been intended by the parties as businessmen.

In general, commercial law is concerned with arm's length transactions creating personal obligations between businessmen. It is not usually thought of in connection with equity and its preoccupation with the trust idea, traditionally in the context of the private family trust of land and other property, and the public trust for the promotion of charitable purposes. So it can be seen why, as Sir Roy Goode recalled in his *Hamlyn Lectures*, it used to be said that 'equity should have no significant role to play in commercial transactions.'[12]

Commercial transactions are mainly governed by commercial law. That expression is a convenient label loosely applied to a ragbag of contract topics relevant to commercial and business transactions. Commercial law is not to be found in any single self-contained code. It consists mainly of the general principles of the law of contract and detailed special common law and statutory rules applying to various kinds of particular contracts and transactions made by commercial men: sale, exchange, agency, negotiable instruments, documents of title, contracts of carriage, insurance, guarantee, and other financial transactions. When disputes arise they are commonly called 'commercial cases' and are determined by the commercial court or in a commercial arbitration.

VII. ECONOMIC LOSS: CONTRACT, TORT AND EQUITY

Commercial disputes are mainly concerned with purely economic loss, for which the usual remedy is compensatory damages. Economic loss was for long considered to be the special province of the law of contract. No contract, no liability for economic loss.

[12] RM Goode, *Commercial Law in the Next Millenium* (London: Sweet & Maxwell, 1998) at p 17.

But the law of contract does not govern the existence of, or exhaust the content of, all consensual relationships. Indeed, it is only in recent years that, apart from the torts of deceit, interference with contractual relations, malicious falsehood, passing off and conspiracy, tort law has begun to play a more significant part in commercial disputes via the tort of negligence and its extension to recovery of purely economic loss for negligent statements and acts, especially in the context of special relationships. In some cases there may be concurrent liability in tort and in contract. In many cases, however, where the parties are already in a contractual relationship, particularly in the case of a commercial contract, there is no particular advantage in making an additional or alternative claim in tort.

Economic loss suffered may even give rise on the same facts to concurrent liabilities in contract, tort and in equity. And even if there is no contract and no tortious wrong, a case of economic loss may give rise to liability for an equitable wrong.

Surprisingly, it is sometimes overlooked that equity has been operating effectively in the area of special relationships and of restitutionary remedies and compensation for economic loss for very much longer than tort has.

Let me cite a recent experience, the case of *Langstaff v Birtles*.[13] The claimants retained the defendant solicitors to act for them on the proposed purchase of a pub called 'The Moorcock.' When the claimants decided not to proceed with that purchase and determined the retainer, the defendant solicitors suggested to the claimants that they might like to buy into the business of a hotel venture in which the solicitors were partners. The claimants agreed. The defendants did not advise the claimants to seek independent advice and they received none. The claimants agreed to become partners in the hotel venture with the solicitors. They paid for their half share. The business failed. The claimants lost everything, including their living accommodation in the hotel. The trial judge rejected the pleaded case of professional negligence, misrepresentation and breach of fiduciary duty as partners. As for the negligence claim, the judge held that there was no claim for breach of contract and that the defendants did not owe any duty of care to the claimants after the termination of the retainer.

As the judge himself appreciated, this result did not accord with the justice of the case, but no attempt was made before him to base the claim on the obvious equitable basis of abuse of trust and confidence by a solicitor, who, in all the circumstances, occupied a position where his duty and his interest conflicted, in selling property to his client or former client, who did not have the benefit of independent legal advice. When the pleadings were amended with the leave of the court on the appeal, it was possible to uphold

[13] [2002] 1 WLR 470.

the claim and award substantial equitable compensation to the claimants for the defendants' breach of duty in equity.

As in tort, the nature and existence of the relationships and the duties arising from them are chiefly determined by the general law and not, as in contract, by the agreement of the parties, though a relevant agreement between the parties may affect the existence and scope of the relationship and the duties as between those parties and others under the general law.

VIII. EQUITY AND RELATIONSHIPS

In general, the focus of equity and the trust idea is more on relationships, which may or may not originate in a contract. In the course of a relationship specific transactions may trigger disputes. For centuries equity has scrutinised the integrity of those relationships in which there is a significant risk of the exploitation of the vulnerable: undue influence and unconscionable bargains, for example, even if the kind of situations in which equity has intervened are not themselves 'commercial', they can have extremely serious consequences for commercial organisations, such as banks, and so become the concern of commercial lawyers and of commercial law.

Take, for example, the impact on banks and bank lending operations of the application of the equitable doctrines of undue influence and of trust and confidence, as between husband and wife, to commonplace security transactions (ie a wife standing surety for and charging her interest in the matrimonial home to a bank to secure her husband's indebtedness): see the decisions of the House of Lords, *Barclay's Bank v O'Brien*[14] and, more recently, *Royal Bank of Scotland v Etridge (No 2)*.[15] As Lord Nicholls explained in his speech,[16] the relevant area in which the courts of equity acted as a court of conscience included not only specific overt acts of improper threats, duress or pressure, but also the prevention of abuse of influence in 'relationship' cases where the sphere of undue influence arose out of a relationship in which one person seeks to take unfair advantage of another who had placed trust and confidence in the other to look after his affairs and interests and over whom he has in the course of a relationship acquired a measure of influence. There is no comprehensive list of the types of relationship caught by this principle and there is 'no single touchstone for determining whether the principle is applicable.' It may even apply in special circumstances to relationships, such as banker and customer, to which it would not normally apply.

[14] [1994] AC 180.
[15] [2002] 2 AC 773.
[16] At paras 6–12.

Lord Nicholls rejected the contention that the application of the doctrine to such cases, though departing from conventional concepts, would have an 'unsettling effect on established principles of contract.'[17] He did not regard the equitable concept of being 'put on inquiry' as imposing more than a modest obligation on banks and other creditors, or as going beyond the reasonable requirements of modern times, so that in future they should regulate their affairs accordingly.[18]

He expressly left on one side the case where the guarantee is given on a commercial basis and 'the relationship between the surety and the debtor is commercial.' He said:

> Different considerations apply where the relationship between the debtor and the guarantor is commercial, as where a guarantor is being paid a fee, or a company is guaranteeing the debts of another company in the same group. Those engaged in business can be regarded as capable of looking after themselves and understanding the risks involved in the giving of guarantees.[19]

In his speech Lord Bingham concisely stated[20] what was required of the law when it sought to apply equitable doctrines of trust and confidence to protect vulnerable wives from abuse of influence and so had an effect on commercial transactions, such as bank loans:

> The law must afford both parties a measure of protection. It cannot prescribe a code which will be proof against error, misunderstanding or mishap. But it can indicate minimum requirements which, if met, will reduce the risk of error, misunderstanding and mishap to an acceptable level. The paramount need in this important field is that the minimum requirements should be clear, simple and practically operable.

Of more direct relevance to commerce is the variety of situations in which fiduciary duties have been recognised and enforced in commercial situations.

Breach of a fiduciary duty is a civil wrong for which restitution of benefits wrongly obtained by the wrong and compensation for loss could be awarded. However, it is only relatively recently that legal advisers and the courts have really started to take an active interest in the considerable potential for the wider application of these aspects of the trust idea in the commercial field and in the rights and remedies available quite apart from those available in the common law of contract and tort.

The contribution made by equity has been increasingly recognised in this country in the influential writings of Sir Roy Goode and of Lord Millett,

[17] Para 43.
[18] Para 89.
[19] Para 88.
[20] Para 2.

judicially, as well as extra-judicially. (The extent of that influence will be evident from the debt that this paper owes to them.)

Time and space do not permit discussion of the equitable doctrines of subrogation and set off or the particular attractions of the range of restitutionary proprietary and personal rights and remedies available in equity or the use of the institution of the trust in commerce.

The pressing problem, as already indicated, is to repair the damage to trust in commercial life and activities, especially in the financial sector. Issues of fiduciary relationships and duties in equity and, to a lesser extent, of confidential relationships and duties are relevant to this problem. They are also at an interesting stage of development.

IX. EQUITY: WHAT DOES IT HAVE TO DO WITH COMMERCE?

In a technical historical sense equity has very little to do with commerce or commercial law. It is the body of rules and principles of substantive law, procedure and practice administered by the Court of Chancery before the fusion of the administration of law and equity in 1875. It is dominated by the institution of the trust which, with its distinctive split of legal and beneficial ownership, was unrecognised by the common law and so was exclusively within the jurisdiction of the Court of Chancery. It was principally concerned with the holding of, the preservation of (from tax, among other things), and the succession to, family wealth, usually in the form of land, and the execution of private and public trusts in which property was held. The trust was not concerned with the enforcement or regulation of arm's length transactions in the market, which increasingly generated much of the wealth used to purchase the land put in trust or with the resolution of commercial disputes between businessmen. As far as equity was concerned, the message was 'Put not your trust in money. Put your money in trust.'

Even in a very non-technical general sense of fairness, evenness, and of doing what is just and equitable on the basis of conscience and reason, equity had little to do with commerce or commercial transactions and disputes. The essential questions in a commercial dispute were whether there was a promise or set of promises constituting a legally binding and enforceable bargain reached between two or more parties, what had been agreed and whether the agreement had been broken. Commerce was not attracted to unsettling doctrines of whether what was agreed was fair or just or equitable or whether anyone in an arm's length transaction negotiated in the market had acted against conscience.

So equity is mainly associated with the institution of the private family trust and of property concepts associated with it, such a beneficial ownership, which had no obvious place in most commercial transactions. This feature has tended to overshadow the pervasive underlying trust idea applicable to relationships outside that of trustee and beneficiary. That trust idea is, however, capable of application to the legal analysis of some commercial relationships and transactions.

In this context equity is used as an additional source of obligations based on flexible principles and general standards of fair dealing applicable to a range of special relationships between people involved in commerce, distinguished by a stronger streak of trust than is to be found in ordinary commercial transactions.

X. FIDUCIARY RELATIONSHIPS

In general a commercial relationship, arising from contract or otherwise, is not a fiduciary relationship. The parties to a commercial relationship are at arm's length from one another, each entering into it in his own interest and being legally entitled, consistently with the due performance of the contract, to look solely after his own interests. The parties may trust one another in an important general kind of way, but their dealings with one another make it clear that they are conducted on a strictly commercial basis of bargain.

Although there are difficulties in arriving at a satisfactory statement of the necessary and sufficient conditions to establish a fiduciary relationship, the gist of it is adequately conveyed in the traditional language of equity as one of trust and confidence. It usually arises in circumstances in which a person has undertaken to hold property for the benefit of another or to act for or on behalf of another in a particular matter or matters. This may be regarded as too general a statement on which to predict the decisions of the courts on what is and what is not a fiduciary relationship. But it is no more general than other statements of legal principle as to what is an agreement or what is the duty of care or what is fraud. The unpredictable element is usually what will be found to be the facts, rather than what is the court's reaction to them.

The classic case is that of a trustee holding property upon an express trust for a beneficiary. This can occur in a commercial situation if that is the effect of what the parties have agreed. They can agree on the coexistence of a contractual relationship and a trust relationship with respect to property, such as a loan. But in the absence of clear agreement or common intention the courts are unlikely to find that an express trustee/beneficiary relationship exists in respect of property in a commercial situation.

Outside the paradigm case of an express trust of ascertainable property for another, something more specific must have occurred than simply putting one's trust in another's trustworthiness in a general sort of way. Something more solid and steady than that is required to establish a fiduciary relationship. Like the trust/beneficiary relationship, which may last for generations, the fiduciary relationship often continues and develops over a longer period than the duration of a commercial contract. In consequence it approximates more to the condition of personal security, ease and continuing commitment generated by friendship than to the ordinary relationships between neighbours or fellow citizens engaging in commerce in the rough and tumble of the market place. There must at least be a specific matter bringing the parties into a state of a closer and more personal commitment than would ordinarily be the case in commerce, leading to a relationship in which one person has put a significant degree of trust in another person, who has so acted that it is reasonable, in all the circumstances, to treat him as having undertaken or assumed responsibility for protecting the interests of that other person rather than his own interests in that matter.

On the application of that principle fiduciary relationships have been held to exist in commercial situations concerning the relationship between principal and agent, between company directors and the company, and between partners, as well as in professional and quasi-commercial situations, such as between a solicitor, an accountant or a broker and their clients.

XI. FIDUCIARY DUTIES

The general nature of fiduciary duties is clear. They are essentially negative in character. The duties do not include the positive and particular transactional obligations which he may be required to perform for the other party to the contract, for example, the supply of financial or advisory services, or the provision of finance or dealing in commodities. The obligations arise from the fiduciary character of his situation within a continuing relationship and relate to ethical standards of conduct which he should observe in his dealings as regards the interests of the other party. The duties are imposed by law to deter the fiduciary from using his position without authority for his own benefit or advantage rather then for the benefit or advantage of the person to whom the relevant duties are owed. If, however, the fiduciary makes full prior disclosure of all the relevant circumstances which might constitute a breach of duty and the person to whom the duty is owed freely gives his informed consent to the actions of the fiduciary, there will be no actionable breach of duty.

The duties are concisely stated in a much cited passage from the judgment of Millett LJ in *Bristol and West Building Society v Mothew*:[21]

> The distinguishing obligation of a fiduciary is the obligation of loyalty. The principal is entitled to the single-minded loyalty of his fiduciary. The core liability

[21] [1998] Ch 1 at p 18A

> has several facets. A fiduciary must act in good faith; he must not make a profit out of his trust; he must not place himself in a position where his duty and his interest may conflict; he may not act for his own benefit or the benefit of a third person without the informed consent of his principal. This is not intended to be an exhaustive list but it is sufficient to indicate the nature of fiduciary obligations. They are the defining characteristics of the fiduciary. As Dr Finn pointed out in his classic work *Fiduciary Obligations*[22], he is not subject to fiduciary obligations because he is a fiduciary; it is because he is subject to them that he is a fiduciary.

The scope and extent of the fiduciary duties owed depends on a careful analysis of all the relevant facts of each particular case. So if there is or has been a contract between the parties, the terms of the contract will be relevant in determining the duties owed. If it is alleged that fiduciary duties were owed and have been breached it will be easier to determine the scope and extent of them if the claims are made against a person occupying a status recognised as giving rise to a fiduciary relationship.

The case will be more fact-intensive if it is sought to establish that fiduciary duties were owed by a person in a commercial situation who does not occupy such a status, but owes fiduciary duties by reason of special circumstances not covered by a contract which has been made or where a contract was never concluded.

XII. PRE-CONTRACTUAL NEGOTIATIONS

A good illustration of the difficulties in laying down clear limits to the role of equity in dealings between commercial men is in abortive negotiations for a contract. Commercial men enter into negotiations for a contract. They reach an informal non-contractual understanding in respect of a contemplated contract or commercial relationship, such as a joint venture, which never materialises. One of the parties to the negotiations then advances against the other a claim in equity. The stock response in commercial law (minus any equitable content) is that, if there was no concluded contract and no tort (such as deceit or negligent misrepresentation), there is no legal liability. Equity should not intervene to create a liability in the place of a contract which the parties never made and which the court has no right to make for or impose upon the parties.

There is much force in that approach. The court should be slow to find non-contractual liability to cover a situation which the parties contemplated would eventually be covered by a concluded contract between them.

[22] PD Finn, *Fiduciary Obligations* (Sydney: Law Book Company, 1977) at p 2.

In some cases, however, the courts have held that one of the parties is bound by an equitable obligation arising out of the negotiations which have not crystallised into a binding contract, for example, where, in the course of the negotiations, a party has given an assurance, or an understanding has been reached, in circumstances that make it inequitable for a party to act inconsistently with it.

In *Pallant v Morgan*,[23] the defendant assured the plaintiff that, if he did not compete with him at a land auction, he would, if he acquired the land, divide it between them. The defendant's bid was successful. He then refused to divide the land with the plaintiff. Equity intervened to prevent the defendant from acting in breach of the assurance. He was compelled to hold the land for their joint benefit, with a direction that it should be sold and the proceeds divided equally after the defendant was repaid the purchase price with interest. The arrangement repudiated by the defendant was not a contract of agency or of partnership or any other kind of binding contract. No actionable misrepresentation or deceit was claimed.

Equity supplied a constructive trust remedy where there was no remedy at common law. Few would quarrel with the justice of the result, though there are disagreements about the juristic basis of the principle. Some text writers treat it as a case of agency. In *Bowstead & Reynolds on Agency*[24] the case is cited for the proposition:

> A agrees with B to bid at auction for property which they both want and promises that if he acquires it he will cede part to B. A holds that part on trust for B and cannot claim to own the whole.

Others treat it as a species of fraud in equity.[25] Leading works on restitution[26] and constructive trusts[27] do not mention it at all.

It may be regarded as an example of the court granting a proprietary remedy against a defendant, as another has placed trust in him in respect of a specific transaction. The circumstances give rise to a fiduciary relationship and to fiduciary duties in respect of that transaction. The defendant acts in breach of fiduciary duty if he seeks to keep for himself a benefit obtained in consequence of the trust placed in him. The proprietary remedy is available, even though there is uncertainty, which would probably be fatal to the contention that there was a concluded contract, about the respective beneficial interests, as the parties to the relationship had failed to agree on the beneficial shares in the land.

[23] [1953] Ch 43.
[24] FMB Reynolds, *Bowstead and Reynolds on Agency* 17th edn, (London: Sweet & Maxwell, 2001) at para 6–110.
[25] RP Meagher, WMC Gummow and JRF Lehane, *Equity: Doctrines and Remedies* 2nd edn, (Sydney: Butterworths, 1984) at para 1209.
[26] R Goff and G Jones, *The Law of Restitution* 6th edn, (London, Sweet & Maxwell: 2002).
[27] AJ Oakley, *Constructive Trusts* 3rd edn, (London: Sweet & Maxwell, 1997).

That decision, and the later case of *Banner Homes Group plc v Luff Developments Limited*,[28] were unsuccessfully relied on by the claimant in the recent case of *London & Regional Investments Limited v TBI plc*.[29] It illustrates the importance of carefully considering the contractual position in the context of an attempt to rely on equity to create proprietary rights on the basis of unconscionable conduct of breach of fiduciary duty.

It was alleged by the claimant that the parties had reached an oral understanding for a joint venture for the development through a jointly owned company of property owned by one of them. The understanding was alleged to have occurred in negotiations between the parties which resulted in a concluded agreement under which the claimant agreed to purchase all the issued share capital of the other's property holding company. The claim was that, as part of that deal, there was an understanding that a joint venture company would be formed to develop the development land, and that, after the share deal had been completed, the defendants acted contrary to the understanding by refusing to proceed with the joint venture for the development of the land. The claimant contended that it was entitled by way of constructive trust to a joint beneficial interest in the development land.

The claim was rejected on the defendants' application for summary judgment against the claimants made on the basis that the trust claim had no real prospect of success. The evidence disclosed that the joint venture, which was the subject of an unresolved conflict of evidence, was expressly stated to be 'subject to contract' in the formal contract for the sale of the shares. The effect of that expression was that the parties intended the joint venture project to remain in a state of negotiation unless and until a formal contract was concluded. In those circumstances there was nothing inequitable in the defendants backing out of the negotiations before a final agreement about it had been concluded or preventing them from refusing to proceed with it. The claimants were not entitled to invoke equity to counter the effect of an express agreement that the joint development venture was 'subject to contract.' It was part of the bargain that the matters remained in a state of negotiation until a future agreement was made. The use of the expression negatived an intention on either side to create any legal obligations in respect of the joint development until a final agreement was concluded. There was no evidence that the parties had subsequently agreed to convert the understanding into an unconditional contract. There was no breach of duty or unconscionable conduct on the part of the defendants in withdrawing from the negotiations and there was no detrimental reliance by the defendants on a revocable non-contractual promise of the

28 [2000] Ch 372.
29 [2002] EWCA Civ 355.

claimants giving rise to a trust of or other proprietary right in respect of the development land.

XIII. CONFIDENTIAL INFORMATION: RELATIONSHIPS AND DUTIES

It is a cliché that this is the information age. It is easier than it has ever been in the whole of human history to generate, gather and disseminate information of every kind on every subject. All commercial organisations have information of value to them and to others. They may wish to protect its value by keeping it private.

Commercial organisations also frequently receive confidential information from others. The information may reveal and provide profitable business opportunities. They may wish to disclose the information to third parties. They may be tempted to use the information for their own benefit or for the benefit of others, as well as, or instead of, for the benefit of those from whom they have received the information.

Control over the disclosure and use of information is frequently achieved by the terms of commercial contracts: employees of commercial organisations are usually subject to express or implied contractual obligations not to disclose or use for the benefit of anyone, other than their employer, confidential commercial information obtained by them during the course of their employment. There are commercial contracts the entire purpose of which is to sell or license the use by another of confidential information. In pre-contractual commercial negotiations a contract may be made to protect the information, which has been disclosed solely for the purpose of the negotiations, from being used for any unauthorised purpose.

Where a contract has not been made, it may be necessary to invoke the equitable doctrines available for the protection of confidential information. A duty of confidence in equity arises when, in the course of a confidential relationship, confidential information is imparted by one person to another for a specific purpose. The purpose may be express or implied from all the circumstances. That duty is broken if the information is used or disclosed for a different unauthorised purpose and without any justification in the public interest.

The main focus of the duty in recent cases has not been in the commercial field at all. In high profile cases ambitious attempts have been made to persuade the courts to develop protection for personal privacy in accordance with Article 8 of the Convention on Human Rights and to protect official secrets from disclosure by ex-members of the intelligence and security services. Different considerations arise in those cases, principally from the conflict with other values of freedom of expression and consideration of the public interest in the workings of government and national security.

It appears, for example, that in the personal privacy cases, such as *A v B*,[30] a different and more generous approach has been taken by the courts to the requirement in equity of a confidential relationship than would be appropriate in a commercial information or trade secret context. It was held that:

> The need for the existence of a confidential relationship should not give rise to problems as to the law. The difficulty will be as to the relevant facts. A duty of confidence will arise whenever the party subject to the duty is in a situation where he knows or ought to know that the other person can reasonably expect his privacy to be protected.[31]

The personal privacy and the state secret cases should not be allowed to overshadow the commercial importance of the duty of confidence in respect of trade secrets and sensitive commercial and financial information, which has not been made the subject of protection by contract.

The recipient of confidential information may also be in a fiduciary relationship, such as agent or as company director, either with the person who has imparted the information to him or with a third party. In those circumstances a breach of fiduciary duty may occur if the recipient of the information makes unauthorised use of the information or of the opportunities presented by exploiting them for the benefit of himself or for the benefit of anyone other than the person to whom he owes the fiduciary duty not to profit from his fiduciary position.

In general, the equitable duty of confidence comes into play when there is either no contract at all, or, if there is a contract, it is silent on the relevant use or disclosure of information. That may occur, for example, after the cessation of the contract and the contract may contain no post-termination provision: this situation has arisen in the case of patent licences during the currency of which confidential information has been supplied to enable the licensees to operate the licence. Disputes may arise as the entitlement of the licensee to use the information after the expiration of the patent and the termination of the licence. The problem may be resolved by implying a term into a contract or, if there is no contract, by use of the duty of confidence in relation to information, which has not yet found its way into the public domain.

XIV. CONCLUSION

In some important respects the problems in the relationship between equity and commercial law are caused by the lack of coherence in the coexistence of common law and equity. This defect in our system was discussed by

[30] [2002] 3 WLR 542; and see also *Campbell v MGN Ltd* [2003] 2 WLR 80.
[31] At para 11(ix).

Professor Andrew Burrows in his inaugural lecture.[32] The distinction between, on the one hand, the common law wrongs of breach of contract and tort, and on the other hand, the equitable wrongs of breach of fiduciary duty and breach of confidence, is confusing. It would be more rational to treat equitable wrongs as breaches of contract or as torts. He suggested, for example, that a more illuminating way of thinking about fiduciary duty is simply as a duty to look after another's interests. Depending on the context, the duty may be to take reasonable care or it may be a stricter duty, such as a duty to make disclosure, to account for benefits received or to avoid conflicts of interest and duty.

A more rational approach to identifying the wrong for which a remedy is sought would lead to more effective application of stricter duties in some areas of commercial law, and a better appreciation of the need for them. There is much to recommend an approach which would achieve that result.

[32] 'We Do This at Common Law But We Do This in Equity' (2002) 22 *Oxford Journal of Legal Studies* 1.

Commentary on 'Commercial Notions and Equitable Potions'

PHILIP R WOOD

I. INTRODUCTION

THE MAIN THEMES of the chapter by Sir John Mummery, 'Commercial Notions and Equitable Potions', are the importance of ethics in commercial life and the way in which the English approach is a development of the trust, the trust as a legal institution, and the concept of commercial trustworthiness. The chapter is a distilled and profound statement of the issues and shows the core values of the English legal system in its formulation of commercial law. I have some brief comments of a general sort which focus on the twin concepts, first of commercial morality and secondly the trust.

A. Micromorality

Many of the issues discussed in the essays in this book are issues of micromorality—that is, they are not basic primitive rules without which societies could not function at all (eg, thou shalt not kill or steal). But micromorality is not nano-morality and in particular micromorality can have macro consequences (eg the demise of large firms and the loss of confidence in markets where the participants abuse the principles). The ability of a jurisdiction to master this micromorality in a fair and predictable manner is a test of its civilisation.

B. Insolvency

Insolvency law is a leading marker of the credentials of a legal system in the field of commercial and financial law (ie whether it is serious about its principles). This is because insolvency causes such devastation in credit economies and sets up great emotions and conflicts about who to protect. As Ernest Rutherford said about physics—'Everything else is stamp-collecting.'

One may take the example of extreme unjust enrichment cases in the form of archetypal moral claims, especially the recovery of the proceeds of embezzlement and of gross violations of commercial morality by fiduciaries. The strength of the principle is tested when the recipient who has to disgorge another's property is insolvent. It is not problematic for the courts to return obvious windfalls held by solvent recipients. But there is a real collision of interests when the law confers super-priority claims on insolvency.

The common law systems answer this challenge by supporting the super-priority claim in cases of abusive takings and allowing a measure, albeit limited, of tracing proceeds through substitutions and mixtures, especially in bank accounts through which misappropriated proceeds must usually travel. In other words, these regimes consider that the protection of property against takings is a higher principle than the protection of the creditors of the miscreant who in their view should not have the windfall representing the proceeds of the victim's assets.

This approach is not generally shared by civil jurisdictions (subject to various exceptions), but not because they do not have the concept of unjust enrichment—they most certainly do. It is because of the comparative absence of proprietary tracing through mixtures on insolvency—the trail runs dry much earlier than in common law systems, and often does not start at all. The result is that the principle does not operate in one of the most important cases where one wants it: cheats often end up in the bankruptcy courts

This example shows how controversial the ethical principles are internationally and how, regardless of the routes by which legal systems got where they are now in historical terms (the influences may not necessarily have been driven by insolvency), the real division of view is nowadays crystallised on insolvency. That is when there is no dodging the issue and when the law has to be at its most ruthless in deciding the winner and the loser.

The relevance of insolvency law is underlined by the fact that in credit economies insolvency is an ever-present risk and one which can strike quite without warning, causing calamitous losses and sometimes threatening the whole economy.

C. Agency Risk

Modern societies are permeated by agency risk—that is, somebody looks after the assets of others, be they directors of a company or a bank, or trustees of a pension fund. It does not matter whether one calls this agency risk or intermediary risk or fiduciary risk—it is the same in function.

The risk which this situation imposes is central to Sir John Mummery's chapter, which gives many telling examples. The separation of ownership and control is fundamental and pervasive in all areas of commercial life. It is also necessary.

In this connection it would be germane to compare the rate of insolvency—which is to some degree a test of prudent management—involving governments and listed corporations (the main economically important guardians of the assets of others in substance) and the rate of insolvency involving individuals personally. The number of countries which have become insolvent over the last 25 years or so is probably around 45 per cent—at least 85 countries, some of them serially insolvent. There are presently around 192 sovereign states. I do not know the rate of insolvency of the total population of listed corporations world-wide over the last 25 years, but one would guess at not more than 10 to 15 per cent—but in any event not anywhere near 40 per cent. As to the rate of insolvency of individuals who have incurred material credit, one would guess at somewhere between 0.5 per cent and seven per cent at the outside. Apart from the country figures, these suggested figures are too crude and unverified to qualify as statistics and there are comparability issues. But nevertheless they are sufficient to draw some conclusions about the measure of the agency risk and where it is most pronounced.

D. Financial Assets

It is a often remarked that the form in which wealth is held has historically shifted from land to intangibles—bank money and marketable investments—and that intangibles form a major sector of our economies. Flows of financial assets are maybe 30 times trade flows. Flows through payment systems are about 30 times world GDP.

Financial assets have to be changed into physical assets in order for them to be used and enjoyed by individuals: hence financial assets are a store of wealth. But it is difficult to assess the present value of an item of this type of property and even more difficult to establish its future value—unlike the depreciation on a car.

The combination of the consequences of intangibility—such as the problems of valuation—plus the need to have others to look after or manage the property—agency or fiduciary problems—places great stress on the proper development of moral codes for those who deal in the property.

This process is made even more exacting by the indelible marks which history has imprinted on attitudes to financial assets. One may note, for example, the shadows still thrown across us by Aristotle's views about money, by the mediaeval attack by the established religions on usury (an early version of 'bank-bashing'), by the impact on France of John Law's disastrous experiments with banking, by the impact of the German inflation in the twenties and the impact of the great 1929 Wall Street crash on the American legal psyche. These resentments run deep and are reinforced by scandals from time to time.

The ability of a legal system to impose a commercial morality without losing its sense of proportion in the face of some scandal or serious abusive event is critical test of a legal system's maturity. Financial assets are tokens—are creatures of our own creation—so they are our servants, not our masters, as is the legal regime which governs their use.

It is noteworthy that there is a discrepancy between the legal and moral regimes applicable to dealing in financial assets compared with goods. For example it is not seen as vicious that a car dealer affirms that 'this little runner is selling like hot cakes' when it is not, but this puffing is treated as a fraud in relation to the underwriting of securities.

To the ordinary moral conscience, it would not seem criminal to pay a subordinated claim before the senior claim—but it is a serious criminal offence for directors in most English common law countries, amongst others (but not the United States): see the criminal prohibition on a company giving financial assistance for the purchase of its own shares

You do not have to verify and file a prospectus when you launch a new line in washing machines. An analyst in the pay of a computer company is not excoriated for puffing a computer.

For various obvious reasons the regime for financial assets has to be stricter. But the degree of discrepancy is an issue.

E. Freedom

English commercial law is exceptionally liberal in the institutions of the law which can be used in the furtherance of transactions, and hence of wealth creation, and hence of survival potential. It is liberal in the sense that the regime is comparatively free of arbitrary and illogical barriers to the structuring of transactions. Any international practitioner is soon made aware of what it means to have a liberal and flexible system in real practical terms.

An example is the use of the trust in commercial transactions: since the device is universally available it can be employed in any context, from mortgage trusts to securitisations, from investment funds to bondholder trusts

This liberality is protected by a high moral tone applicable to commercial dealings, a moral tone standards of which are rigorously enforced by the courts. For example, it was no surprise when some years ago the Law Commission advised that the common law on fiduciary duties developed by case law were tougher than those set out in the regulatory codes of conduct applicable to financial services.

These moral standards are stringent but not doctrinaire. A good example of the balance within the system is how it deals with the endemic problems of conflicts of interest which inevitably spring from the prevalence of intermediaries and 'agents' in relation to financial assets. The case law on Chinese Walls shows how the courts cope with the problems in a sensible way which recognises both the moral demands and also the context in which they are played out.

F. The Trust

Finally one may remark on the institution of the trust which is discussed in Sir John Mummery's chapter.

I am not concerned with the minority use, economically speaking, of the trust for testamentary dispositions and family settlements. The use of the trust in the commercial arena massively outdistances these applications: the amounts are spectacular. It seems difficult for modern economies to do without the trust in the operation of commercial life.

Common lawyers often define the trust in terms of fiduciary obligations owed by trustee to beneficiary. They take it for granted that it is effective on insolvency and the cases which establish this date back to the seventeenth century at least.

The essence of the trust is that the trust is a separate patrimony of assets immune from the trustee's private creditors. Yet this position is in theory not accepted by more than half of the world's 300 jurisdictions. In their view the trust assets go to the trustee's private creditors on the trustee's insolvency so that the trust is virtually useless. This approach arose for reasons influenced by the polarised historical development of legal systems up to the 19th century with the added push of diverse attitudes in that crucial period when commercial legal systems were crystallising.

Despite patchwork a of statutes in non-trust countries of establishing settlement systems which are basically trusts (one of these in Belgium holds more securities than the GDP of France), there are still significant mismatches and startling disagreements in the world on basic elements of the legal infrastructure. In that respect the enactment by China of a trust law in 2001 is perhaps one of the most important legal events in world law in the last 25 years.

G. International Role of English Law

English law is widely used for important contracts, especially in the financial arena: consider syndicated credits, bond issues and the ISDA master agreement. The main other widely used system is New York law, and between them they have a near monopoly in some sectors—they are the brewery-horses of governing law.

This happened for many reasons, most of which were not associated with the content of the legal systems concerned, such as economic hegemony, language, location of commercial and financial centres and the like. It did not happen because these systems were more 'moral' than the others, because that is not true.

There are many aspects which international contractors value in a legal system. But there can be no question that contractors will not trust and use a legal system unless its content and the enforcement of the content by the courts represent, in moral terms, a race for the top coupled with a hand on the brake when the cart is going over the top.

Any legal system which seeks to serve a wider role than limited application within its own borders might be thought to face an irresolvable conflict between international use and national use. Clearly the intensity of the protective regime differs according to whether the participants are retail or wholesale, but, apart from this tiering which is a feature of all advanced systems, it is difficult to see why there should ever be a tension between national and international use, since commercial morality in legal matters is not divisible. The view that somehow the ethical content of legal systems must differ because cultures are different can be vastly overstated.

II. CONCLUSION

There is a view that legal micromorality does not matter much: what is important rather is that the jurisdiction has basic standards and, even more importantly, has the resources and expertise to enforce them in an impartial and predictable way. The rest, on this view, is metaphysics.

My view is that the micromorality is crucial and has to be carefully and energetically worked out. It is not simply a matter for one's sense of what is just: it is obvious that commercial and financial markets can be, and often are, seriously compromised by non-compliance with standards of conduct which in their refined sophistication are light-years away from the basic criminal law.

I also suggest that the other strand alluded to in the paper on which I am commenting—the institute of the trust—is a symbol or metaphor of another crucial requirement of modern legal systems: that they should be as free as is consistent with manifest policy restriction. Commercial law will not be able to do its job without this liberality.

3

Statutory Ingredients in Common Law Change: Issues in the Development of Agency Doctrine

DEBORAH A DEMOTT*

I. INTRODUCTION

THIS CHAPTER ANALYSES the roles that statutes may play in the development of common-law doctrine. I focus on developments in the law of agency and primarily but not exclusively on material from the United States. A relationship of common-law agency exists when one person, the 'principal,' manifests assent to another person, the 'agent,' that the agent 'shall act on the principal's behalf and subject to the principal's control, and the agent manifests assent or otherwise consents so to act.'[1] The law of agency thus encompasses a wide cast of characters and range of circumstances in which one person's actions have consequences for another person's legal relations. Agency's large cast of characters includes transactional intermediaries, such as stockbrokers and real-estate agents, as well as individuals empowered to act on behalf of organisations and other non-individual persons, such as corporations. Given its broad reach, it is not surprising that agency raises many kinds of questions concerning connections between the common law and statutes.

The chapter begins with a general sketch of the range of relationships in contemporary US law between common-law agency doctrines and statutes and regulatory rules. It is evident that the contemporary context in which common-law agency operates is one in which statutes of all sorts are

* David F Cavers Professor of Law, Duke University School of Law. Since late 1995, I have served as the Reporter for the Restatement Third, Agency. The opinions expressed in this paper are my own, not those of the American Law Institute, which sponsors Restatements. For comments on earlier drafts, I thank James Penner, Tom Telfer, and participants in the LSE seminar on Commercial Law and Commercial Practice, as well as participants in a faculty workshop at the University of Arkansas, Leflar Law Center.

[1] Restatement Third, Agency § 1.01 (Tentative Draft No 2, 2001).

significant, in which statutes play varied roles, and in which the relationships between statutes and the common law generate hard as well as easy questions. Within this context, I argue that common-law doctrines often supply a degree of structural and definitional coherence amidst a profusion of statutory material. Notwithstanding the practical and intellectual importance of enhancing coherence within the law, how best to achieve coherence in specific instances can be open to question. The paper turns next to the implications of a general and long-running debate concerning the relative merit of statutes and the common law as sources of general normative principles. At least when the focus is agency, the terms of this debate miss the point that often the relationship between statutes and common-law doctrine is one of mutual exchange and influence, as opposed to exclusivity or antagonism. The debate is also characterised by heavy usage of metaphors, which may prejudge or distract attention from substantive questions. The paper concludes with a series of concrete instances, drawn from agency, that illustrate different bases on which to evaluate whether statutory developments should shape common-law doctrine.

II. RELATIONSHIPS BETWEEN STATUTES AND COMMON-LAW DOCTRINES

A. Sources of General Law of Agency

Mapping the landscape of relationships between statutes and common-law agency doctrines is a more complex topographical exercise than one might initially expect. This section sketches a few such relationships. For starters, regardless of the topic, federalism shapes many dimensions of the legal landscape in the United States. On many questions, state law governs. Although the states' law of agency is often characterised as 'common law,' this may not be entirely so. In seven states, legislation codifies many general doctrines of agency law. In six of these states—Alabama,[2] California,[3] Georgia,[4] Montana,[5] North Dakota,[6] and South Dakota[7]—the codifications date from the mid-nineteenth century through the early

[2] See Ala Code § 8-2-1 to -9 (1993) (initial codification 1923).
[3] See Cal Civ Code §§ 2019 to -22; 2026 to -30; 2295 to -2300; 2304 to -39; 2342 to -45; 2349 to -51; 2355 to -57 (1985 & 1999 Supp) (initial codification 1872).
[4] See Ga Code §§ 10-6-1 to -6; 10-6-20 to -39; 10-6-51) 0 to -64; 10-6-80 to -89; 10-6-100 to -102; 10-6-120 to -21 (1996 & 1999 Supp) (initial codification 1863).
[5] See Mont Code Ann. §§ 28-10-101 to -105; 28-10-201 to -215; 28-10-301 to -303; 28-10-401 to -423; 28-10-501 to -503; 28-10-601 to -609; 28-10-701 to -704; 28-10-801 to -802 (1999) (initial codification 1895).
[6] See ND Cent Code §§ 3-01-01 to -11; 3-02-01 to -16; 3-03-01 to -09; 3-04-01 to -03; 3-05-01 to -02; 3-06-01 to -06 (1975 & 1999 Supp) (initial codification 1877).
[7] See SD Codif Laws ch 59-1 to -9; 59-2-1 to -7; 59-3-1- to -18; 59-4-1 to -2; 59-5-1- to -3; 59-6-1 to -10; 59-7-1 to -8; 59-8-1 to -2; 59-9-1 to -8 (1993 & 1999 Supp) (initial codification 1877).

twentieth century.[8] Their content reflects the common law of agency as understood at the time of codification. The codifications, like many common-law doctrines in agency, are formulated in general language and use concepts that are dependent on factual context. Outcomes reached when these codifications apply often do not differ from outcomes reached by courts applying the common law.[9]

In the seventh 'Code' state, Louisiana, the codified treatment of agency originated with provisions based on the French Civil Code. The current codification of agency doctrine, enacted in 1997, is distinctive and draws in some respects from the current French code as well as the Quebec Civil Code.[10] One measure of the distinctiveness of Louisiana's current codified treatment of agency is that it embraces the doctrine of the undisclosed principal.[11] The doctrine—a hallmark of the common law of agency—treats an agent's

[8] The codifications in California, Montana, North Dakota, and South Dakota were based on the Field Civil Code. On the codification movement in the nineteenth-century United States, see Lawrence M Friedman, *A History of American Law* (New York, Simon and Schuster, 1973) 340–55; Mattias Reimann, 'The Historical School Against Codification: Savigny, Carter and the Defeat of the New York Civil Code', 37 *Am J Comp L* 95.

[9] When this is not so, the divergence may simply reflect the passage of time since legislative adoption of the code when counterpart common-law doctrine has changed in the interim. For example, under Cal Civil Code § 2335, when a third party has dealt exclusively with an agent for an undisclosed principal, the principal's liability to the third party is discharged by the principal's payment to the agent if made in good faith and before notice that the third-party creditor elects to hold the principal liable. See *Shasta Livestock Auction Yard, Inc v Bill Evans Cattle Management Corp*, 375 F Supp 1027 (D Idaho 1974) (applying Cal Civ Code § 2335). In contrast, most cases decided after 1958 follow the rule as stated (in 1958) in Restatement Second, Agency § 208, which does not treat good-faith payment by an undisclosed principal to the agent as a basis for discharging the principal's liability to the third party. Cases adopting the rule as stated in § 208 include *Poretta v Superior Dowel Co*, 137 A2d 361 (Me 1957) and *A Gay Jenson Farms Co v Cargill*, Inc, 309 NW2d 285 (Minn 1981). Most US cases decided prior to 1958 followed the rule of good-faith payment codified in Cal Civil Code § 2335. See Restatement Second, Agency § 208, Reporter's Notes (1958).

The underlying rationale for the position taken in Restatement Second § 208 is that an undisclosed principal owes a duty of performance to the third party that is not discharged by rendering performance to someone else, including the agent. The Restatement rule is also justifiable on pragmatic grounds because, under the doctrine codified in Cal Civil Code § 2335, the dispositive question will often be whether the principal acted in good faith in settling with the agent. Section 2335 places on the third party, who did not observe interactions between a principal and an agent, the burden of establishing that the principal knew or had reason to know that the agent was insolvent or otherwise unreliable, as well as requiring the court to draw fine shadings of culpability that depend on whether the principal may have had a basis for qualms about the agent. See also *Poretta*, 137 A2d at 373 (emphasizing the value of adopting a rule that is 'clear cut and explicit').

[10] See La Civ Code arts 2985–3032 (Above 1999) (successor to Civil Code of 1870).

[11] See La Civ Code art 3023 ('[a] third person with whom a mandatary contracts without disclosing his status or the identity of the principal is bound to the principal for the performance of the contract unless the obligation is strictly personal of the right non-assignable.'). Art 3020 provides that a principal, whether disclosed or undisclosed, is 'bound to perform the contract that the mandatary, acting within the limits of his authority, makes with a third person.' Prior to the 1997 legislation, Louisiana cases recognized the possibility that an undisclosed principal may be subject to liability on a contract made on the principal's behalf by an agent. See *Woodlawn Park Ltd P'ship v Doster Constr Co*, 623 So 2d 645, 648 (La 1993).

transaction with a third party as effective to create rights and obligations in the principal on whose behalf the agent acted, although the third party had no reason to think the 'agent' acted on behalf of anyone else. In contrast, civilian systems of agency law 'tend to draw a bright line between acting in the name of another and acting in one's own name ....'[12] As a consequence, within the civilian tradition an agent who deals with a third party exclusively in the agent's own name does not affect legal relations between the third party and the agent's undisclosed principal.[13] Louisiana's position infuses, into a codified civilian scheme, one of the distinctive elements of common-law agency doctrine.

B. Common-Law Definitions of Statutory Terms

The fact that a statute applies to a particular situation does not exclude the possibility that common-law doctrine may be applicable as well. For example, within a context defined or primarily governed by a statute, common-law doctrines may define the meaning of terms used in the statute. Common-law doctrines may also supplement the statute's operative consequences. Common-law agency doctrines form part of the content of many statutory schemes because applying the statute requires reference to definitions or doctrines drawn from the common law. Frequently a statute or an administrative regulation uses terminology from common-law agency—such as 'agent'—but provides no independent definition of the term. Some prominent statutes expressly refer to the common law of agency to define a term used in the statute. For example, Internal Revenue Code § 3121(d) defines 'employee' as 'any individual who, under the usual common law rules applicable in determining the employer-employee relationship, has the status of an employee.'

In contrast, when a statute uses a term defined by common-law doctrine but does not expressly incorporate the doctrine, questions may arise about whether or to what extent the legislature intended incorporation. It may be open to question whether statutory usage of a term such as 'agent' should

The Code distinguishes between two different forms of agency, procuration and mandate. Procuration is a 'unilateral juridical act' by which one person confers authority on another. La Civ Code art 2987. A mandate is 'a contract by which a person, the principal, confers authority on another person, the mandatary, to transact one or more affairs for the principal.' *Ibid*, Art 2989.

[12] See Konrad Zweigert & Hein Kötz, *Introduction to Comparative Law* 3rd edn (New York, Oxford University Press, Tony Weir transl 1998) 433.

[13] The differences may be less as a practical matter than theory would suggest because Continental systems recognise a concept of indirect representation, under which an agent deals in the agent's own name but on behalf of a principal. The principal may able to sue on contracts made by the agent and to intervene in the agent's bankruptcy. However, indirect representation does not usually subject the principal to liability to third parties. See Francis MB Reynolds, *Bowstead & Reynolds on Agency* 17th edn (London, Sweet & Maxwell, 2001) para 1–021, at 10.

be understood literally and whether such usage should invoke all its common-law consequences. Some judicial statements suggest that statutory usage of a term freighted with common-law consequences automatically triggers them, unless the statute provides otherwise. Recently, in construing the language of a statute prohibiting discrimination in employment, the US Supreme Court endorsed the proposition that '[w]here Congress borrows terms of art in which are accumulated the legal tradition and meaning of centuries of practice, it presumably knows and adopts the cluster of ideas that were attached to each borrowed word in the body of learning from which it was taken and the meaning its use will convey to the judicial mind unless otherwise instructed.'[14]

Sometimes, however, a court may conclude that agency terminology in a statute should not be taken literally. For example, the California corporation statute requires a corporation to indemnify any person who is or is threatened to be made a party to litigation against the costs of a successful defense when the person's status as an actual or threatened party is 'by reason of the fact that the person is or was an agent of the corporation.'[15] The same section of the statute defines an 'agent' as 'a director, officer, employee, or other agent.'[16] In *Channel Lumber Co v Porter Simon*, the party seeking indemnity was the corporation's outside litigation counsel who prevailed in a malpractice action brought by the corporate client.[17] The court denied the indemnity, limiting the scope of 'agent' entitled by the statute to indemnity to persons who may act on a corporation's behalf and in its place, as in transactions.[18] Read more broadly, the statute would permit fee-shifting by lawyers embroiled in litigation with their corporate clients, in the absence of any specific indication in the language of the statute that the legislature intended to create such an exception to the general American doctrine barring fee-shifting. Moreover, a broad reading of the statute would not further the purpose generally ascribed to indemnification provisions in corporation statutes, which is to make service as an officer or director more attractive by reducing the risk of dauntingly high litigation costs. Reducing this risk may also benefit a corporation's shareholders, who will not benefit if directors are unwilling to make business decisions that carry some risk.[19] In contrast, the risks of business decision-making would

[14] *Kolstad v American Dental Ass'n*, 119 S Ct 2118, 2126 (1999) (quoting *Morissette v United States*, 342 US 246, 263 (1952)). The question addressed in *Kolstad* is the meaning to be ascribed to a statute's usage of the terms 'reckless indifference' and 'malice' in the context of whether punitive damages are available to a plaintiff. In *Morissette*, the question was whether criminal intent was a necessary element of a statutorily-defined crime of embezzling, stealing, purloining, or knowingly concealing government property.

[15] See Cal Corp Code § 317(c).

[16] See *ibid*, § 317(a).

[17] *Channel Lumber Co v Porter Simon*, 93 Cal Rptr 2d 482 (App Ct 2000).

[18] *Ibid*, at 488–90.

[19] For the justification for indemnification provisions, see Model Business Corporation Act, Subchapter E, Introductory Comment (noting consensus that directors should have 'appropriate

not be reduced by requiring a corporate client to indemnify its former counsel against costs incurred in defending malpractice litigation.

More obliquely, a statute may invite courts to turn to the common law to define a term because, although the statute provides an explicit internal definition of a term, it is circular. For example, two prominent federal statutes—the Employee Retirement Income Security Act ('ERISA') and the Age Discrimination in Employment Act ('ADEA')—both define an 'employee' as an 'individual employed by' an employer.'[20] In *Nationwide Mutual Ins Co v Darden*, the Supreme Court turned to the 'general common law of agency' to supply the operative content requisite to applying ERISA's definition of 'employee,' which requires a multi-factored examination of whether the 'employer' has the right to control the manner and means by which the 'employee' accomplishes a project.[21] In theory, the Court's turn to common-law interpolation was not strictly necessary because the Court could have developed an ERISA-specific definition of 'employee' without borrowing from common-law agency. But the statute itself provides no clues for how such a definition should be developed and, in the absence of reason to believe that the definition independent of common-law content was necessary to fulfill the legislative intention underlying ERISA, it is not surprising that the Court did not undertake this exercise.

Some statutes also refer to concepts that are defined by common-law agency doctrines. For example, under UCC § 1–203(43), an 'unauthorized' signature is defined as 'one made without actual, implied, or apparent authority,' including a forgery. The Code thus refers to common-law doctrine to determine whether a person is bound by a particular signature. Moreover, § 1–103 makes the common law of agency, as well as other general bodies of law, applicable unless specifically displaced by particular

protection against personal risk,' including costs of defense). On benefit to shareholders, see Principles of Corporate Governance: Analysis and Recommendations § 7.19, Comment *c*, at 240 (1994) ('the threat of liability may make corporate officials excessively risk-averse in their decisionmaking, thereby injuring shareholders and diminishing efficiency.').

[20] See 29 USC § 1002(6) (for ERISA purposes, an employee is 'any individual employed by an employer'); 29 USC § 630(f) (for ADEA, purposes, an employee 'means an individual employed by any employer').

[21] See 503 US 318, 323 (1992) ('[i]n determining whether a hired party is an employee under the general common law of agency, we consider the hiring party's right to control the manner and means by which the project is accomplished. Among the other factors relevant to this inquiry are the skill required; the source of the instrumentalities and tools; the location of the work; the duration of the relationship between the parties; whether the hiring party has the right to assign additional projects to the hired party; the extent of the hired party's discretion over when and how long to work; the method of payment; the hired party's role in hiring and paying assistants; whether the work is part of the regular business of the hiring party; whether the hiring party is in business; the provision of employee benefits; and the tax treatment of the hired party … all of the incidents of the relationship must be assessed and weighed with no one factor being decisive.'). See also Restatement Second, Agency § 220(2) (1958) (stating non-exclusive list of factors relevant to whether a person acts for another as a 'servant' or an independent contractor).

provisions in the UCC. Indeed, the Code's drafters contemplated a two-way flow of influence between the statute and the common law by specifically endorsing extension of the applicability of Code doctrines to analogous circumstances that the Code does not itself cover.[22]

C. Common-Law Bridges

Common-law agency doctrines may also serve as bridges between statutes. Bridging becomes essential when one statute, such as the UCC, specifies the operative legal consequences of actions taken by persons and another statute, such as a corporation or other organisational statute, enables private parties to create a legally-constituted person with distinct legal personality. By providing doctrines such as actual and apparent authority, agency bridges the consequences of the organisational statute with legislation (and common-law doctrines) that specify the legal consequences for persons of actions that they take or neglect to take. Bridging is an important element in the overall functionality and coherence of any legal system with even a moderate amount of complexity.

A recent illustration of the functional significance of common-law bridging is the Supreme Court's opinion in *Meyer v Holley*, a private action under the federal Fair Housing Act.[23] The Act, which forbids racial discrimination by 'any person' with respect to the sale or rental of a dwelling, defines persons to include corporations and other organizations but does not address the question of vicarious liability.[24] The Court, beginning with the premise that the Act should be read against the common-law background of 'ordinary tort-related vicarious liability rules' in the absence of any indication that Congress intended otherwise, held that vicarious liability for an individual's violation of the act extended to the corporation that employed the individual but not to the corporation's sole shareholder and officer.[25] Common-law agency creates vicarious liability for employers and other principals, but not for superior agents within an organisational hierarchy. Common-law agency also distinguishes between a corporate principal and its shareholders and does not conflate them.

Common-law doctrines also serve bridging functions when they provide supplemental content that is complementary to a statute. Contemporary

[22] UCC § 1–102, Official Comment 1. See also E Allen Farnsworth, *Contracts* 3rd edn (New York, Aspen Law & Business, 1999) 39; Note, 'The Uniform Commercial Code as a Premise for Judicial Reasoning', 65 *Colum Law Review* 880. To the same effect is § 57(1) of the Sale of Goods Act in Ontario, RSO 1990.

[23] 123 S Ct 824 (2003).

[24] 42 USC § 3605(a) (prohibition of discrimination by any person 'whose business includes engaging in residential real estate-related transactions') and § 3602(d) (defining '[p]erson' to include corporations and other organisations).

[25] 123 S Ct at 828.

legislation creates regulatory structures and authorises the administrative prescription of regulatory rules applicable to many sorts of agents but does not preclude the applicability of common-law agency doctrines to resolve questions not resolved by the legislation. Even complex and extensive regulatory schemes are not likely to be completely comprehensive in providing bases to resolve all questions that may arise. For example, the relationship between a stockbroker and the broker's customer is one of common-law agency when the broker buys or sells on the customer's behalf. In the United States, a duties of a broker in exchange-traded securities are prescribed by the Securities and Exchange Commission in rules adopted pursuant to rule-making authority granted by federal securities legislation.[26] A broker's duties are further prescribed by the National Association of Securities Dealers, a self-regulatory organisation authorised by the SEC

However, even amidst such detailed and complex regulation, common-law agency doctrines apply to the broker-customer relationships and to broker-customer disputes not governed by the statutory and regulatory scheme. For example, in *Merrill Lynch, Pierce, Fenner & Smith, Inc v Cheng*, a broker executed an unauthorised option transaction on behalf of a customer.[27] The court applied common-law agency doctrine to assess whether the broker sufficiently disclosed the unauthorised activity and its consequences to the customer, including the customer's right of rescission. Common-law agency, that is, may serve an interstitial function when a regulatory scheme is less than completely exhaustive in how it addresses the legal consequences of the relationships it covers.

In other respects, however, recent cases suggest the lack of a comprehensive judicial theory of the relationship between common-law doctrines in the context of federal securities regulation. The implications of this point overrun the modest bounds of this paper. A well-known instance is *Central Bank of Denver, NA v First Interstate Bank, NA*, in which the Supreme Court eliminated aiding and abetting as a basis for liability in federal securities fraud actions brought by private plaintiffs; the Court's rationale was the statute itself does not expressly create such liability.[28] However, it has long been recognised that a corporation itself may be subject to the legal consequences created by the securities laws on the basis of statements made by its officers or directors, although the securities laws do not expressly

[26] See, eg, Securities Exchange Act § 15(a) (requiring registration with SEC of any broker or dealer in securities, unless business is 'exclusively intrastate and … does not make use of any facility of a national securities exchange') and § 15(f) (requiring registered broker or dealer to establish, maintain, and enforce rules to prevent misuse of nonpublic information by broker or dealer, or by persons associated with it, in conformity with rules adopted by SEC). See generally Louis Loss & Joel Seligman, *VII Securities Regulation* 3rd ed (Boston, Little Brown, 1989) 3107–3328.

[27] See *Merrill Lynch, Pierce, Fenner & Smith, Inc v Cheng*, 901 F2d 1124 (DC Cir 1990).

[28] 114 S Ct 1439 (1994).

so state.[29] To be sure, the outcome in *Central Bank* is not necessarily inconsistent with attributing legal consequences to corporate principals. *Central Bank* might represent a minimalist position on incorporation of common-law doctrines that treats them as operative only when provisions in the relevant statute would otherwise be inoperative. The statutory imposition of liability for fraud on 'any person' is language that does not apply to a legally-created person, such as a corporation, unless acts of natural persons may be attributed to it. The Court's opinion, however, does not explore these underlying questions.

D. Forms and Limits of Coherence

Interrelationships among common-law doctrines and statutes contribute coherence to contemporary law by providing consistency in meaning across distinct bodies of law. Without mechanisms to develop common grounds of reference, as ever more legislation comes into force, there is a risk that more and more of the law as a whole will consist of insular instances defined by vocabularies of terms and concepts that resist translation into other contexts.[30] Intellectual fragmentation has many consequences. These include a higher degree of specialization within the legal profession, in which each cohort deploys its own specialized vocabulary, displaying at its extreme an 'aura of pontifical esoteric professionalism' as Otto Kahn-Freund termed it.[31] Within a codified legal system, the code provides a counterweight against fragmentation of meaning. Within common-law systems, common-law concepts provide a legal alphabet of sorts,[32] a manageable number of

[29] See, eg, *In re Atlantic Fin Management, Inc*, 784 F2d 29 (1st Cir 1986). See generally James D Cox, Robert W Hillman, & Donald C Langevoort, *Securities Regulation* 2nd edn (New York, Aspen Law & Business, 1997) 696–7 (noting inconsistency and that 'many courts seem to assume that corporations and other business organizations can be primary violators [of statute] without resort to any secondary liability theories;' given that basic prohibition on fraud in Securities Exchange Act of 1934 , section 10(b), refers to 'fraud by any person' and the '34 Act defines person to include corporations and other entities, some have reasoned that Congress intended for there to be direct corporate liability when false or misleading statements were made on the company's behalf by officials acting within the scope of their authority.').

[30] For an account of another source of lack of overall coherence in the law, see William M C Gummow, *Change and Continuity: Statute, Equity, and Federalism* (New York, Oxford University Press, 1999) 26 (arguing that 'a wilderness of single instances' will result as a consequence of statutes that resolve difficult issues by providing for the exercise of discretion by a court or an administrative decision-maker).

[31] Otto Kahn-Freund, 'Reflections on Legal Education', 29 *Mod L Rev* 121, 131, reprinted in Otto Kahn-Freund, *Selected Writings* (London, Stevens & Sons, 1978) 359, 368. Thanks to Simon Deakin for this reference.

[32] See *ibid*. Sir Otto Kahn-Freund credits the 'legal alphabet' phrase to Jhering's work on Roman law. Jhering contrasted the structure of Roman law, which reduced law to a relatively small number logical components, with ideogrammatic Chinese script, in which each concept has a special sign and 'new concepts necessitate in the first place the making of the appropriate new signs. We however have a small alphabet, by means of which we can dissect and compose every word.' *Ibid*.

terms carrying operative consequences that can be applied in diverse statutory contexts.

The examples discussed in this chapter illustrate that consistency or coherence in meaning among statutes and common-law can have two distinct forms that are not mutually inconsistent. What may be termed 'vertical' consistency is present when the same term has the same meaning when applied in a statutory scheme as in the common law. The term 'employee,' for example, is vertically consistent in meaning in the common law of agency and in the statutory context of ERISA as a consequence of the Supreme Court's opinion in *Darden*. What may be termed 'horizontal' consistency is present when the same term is consistent in meaning across multiple statutes, for example if 'employee' has the same operative meaning across multiple statutes. 'Employee' would be horizontally as well as vertically consistent in meaning to the degree courts use the same common-law definition in applying statutes that use the term.

The extent to which courts should achieve vertical and horizontal consistency in applying federal employment statutes is a structural issue that underlies several current disputes. In *EEOC v Sidley Austin Brown & Wood*, the majority of a Seventh Circuit panel held that, for ADEA purposes, detailed factual investigation would be necessary to determine whether individuals termed 'partners' by a law firm were instead 'employees' whose demotion might contravene the ADEA.[33] An important if not decisive question in this inquiry will be the degree to which profits are spread unevenly across the entire firm. The partnership agreement delegated 'absolute power' to the firm's executive committee; although the demoted partners were personally liable for the firm's debts, a high concentration of profits in the firm's executive committee 'would bear on the unavoidably multi-factored determination of whether this large law firm … should for purposes of antidiscrimination law be deemed the employer of some at least of the individuals whom it designates as partners.'[34] The concurring opinion agrees that the firm's designation is not dispositive.[35] However, it argues that, to reduce uncertainty for courts and large professional partnerships, the court should be clear that an individual who is a 'partner' for other purposes should also be characterised as a partner and thus not an 'employee' for ADEA purposes.[36] The concurring opinion also notes the high degree of similarity with which federal employment legislation defines 'employee.'[37]

What the concurring opinion in *Sidley Austin* champions is vertical consistency in meaning between federal employment law and terms defined by the

[33] 315 F3d 696 (7th Cir, 2002).
[34] *Ibid*, at 706.
[35] See *ibid*, at 708.
[36] *Ibid*, at 710.
[37] *Ibid*, at 708.

common law and organizational law: 'We should ask not what the organization, or any given state, *calls* this person; we should ask how this set of attributes is classified under the prevailing law of agency.'[38] Implicitly, the opinion supports horizontal consistency as well based on common language in among federal statutes.[39] The majority opinion, in contrast, implies that for each statutory context—and arguably for each large partnership whose characterisations are in dispute—the court should weigh multiple factors to determine how a firm should be characterised, based on the functional reality of how the firm is run and how it divides its profits. The same structural question also underlies the disarray among federal circuit courts in determining whether a small firm's decision to incorporate should be disregarded for purposes of federal employment law; if a court recharacterises shareholder-employees of an incorporated firm as partners, the firm may have too few employees to trigger the applicability of federal employment legislation.[40]

However, achieving consistency is not a satisfying accomplishment when it requires ignoring underlying differences in meaning and operative consequences. The 'legal alphabet' approach may be misleading if courts implement it by simply transplanting, into a statutory context, the results of a formal exercise in ascertaining how a person or a relationship would be classified by common-law doctrine or under an organisational statute. The 'legal alphabet' metaphor may mislead by distracting attention from the purpose of the statute, as well as from the context in which a particular common-law doctrine does its work.

Consider in this connection the distinct ways in which a person may become subject to liability as a partner. Partnership, like agency, is a consensual relationship; one person becomes another's partner through an association as coowners of a profit-seeking business.[41] In contrast, a person becomes a partner by estoppel when the person falsely represents, or permits others to represent, that the person is a member of a partnership. A partner

[38] *Ibid*, at 708 (emphasis in original).
[39] See also *Hyland v New Haven Radiology Assocs*, PC, 794 F2d 793, 795 (2nd Cir 1986) (given similar purpose—'to stamp out discrimination in various forms'—of ADEA, Federal Labor Standards Act, and Title VII of Civil Rights Act of 1964, courts accord persuasive authority to interpretation of one statute's definitional provisions in construing similar provision in another statute).
[40] Compare *EEOC v Dowd & Dowd, Ltd*, 736 F2d 1177, 1178 (7th Cir 1984) (federal court must assess for itself whether 'economic realities' establish existence of employment relationship) with *Clackamas Gastroenterology Assocs, PC v Wells* 271 F3d 903, 905–06 (9th Cir 2001), cert granted, (US 1 October 2002) (although classification as a 'partner' or 'independent contractor' is never conclusive in an employer's favor, firm that elects to incorporate is bound by consequence of its choice for purposes of applicability of federal employment legislation). The Supreme Court resolved this question by reversing *Clackamas* and abrogating *Dowd & Dowd*. See 123 S Ct 1673 (2003). The Court held that this question should be resolved, as in *Darden*, by applying the common-law definition of 'employee'.
[41] Unif Partnership Act § 6(1) (1914); Rev Unif Partnership Act § 202(a) (1996).

by estoppel is subject to liability to a third person who, relying on the representation of partnership, enters into a transaction with what is or appears to be a partnership, and, if the members of the actual or apparent partnership consent to the representation, the person who makes it acts as their agent.[42] Although a partner by estoppel is subject to liability to third parties who rely on the representation of partnership, and although those who consent to the representation become subject to liability to third parties as if a partnership existed, a partner by estoppel is not otherwise treated as a partner. Thus, a 'partner' by estoppel is not in a fiduciary relationship with the persons he falsely represents to be his partners, unlike a person who is a 'partner' through association with others as coowners of a business. Similarly, if a person falsely purports to be another person's agent and enters into a transaction with a third party purportedly on the principal's behalf, the purported agent breaches an implied warranty of authority that he acts with power to bind the purported principal, just as a true agent would breach the agent's warranty of authority if the agent lacked power to bind the principal to a transaction with a third party.[43] A purported agent is not otherwise treated as an agent. In short, the operative consequences associated with 'partner' and 'agent' are not uniform; they depend on why a person is so characterised and how the question arises.[44]

Thus, the alphabet of terms provided by common-law and organisational law doctrines is more complex than it may appear on first glance. The same term—'partner'—does not always connote the same consequences. If the common law and organisational law may be understood as creating an alphabet of terms, the alphabet is analogous to one for a language—like French or German—that uses diacritical marks that take into account how or where a particular letter is used. The common-law exercise is not a purely formal one of classification, detached of context. An implication is that a statute's usage of a term that connotes consequences defined by common-law doctrines should not be dispositive of how a court should define the term in the context of the statute. The purpose of the statute is relevant, just as the context or structure of a dispute governed by common-law doctrine shapes how doctrine is formulated and the point of the formulation.

III. CONTESTED TURF BETWEEN COMMON-LAW AND STATUTES

The relative merit of statutes and the common law as sources of normative principle is a long-lived controversy. This debate has not much focused on

[42] Unif Partnership Act § 16(1) (1914); Rev Unif Partnership Act § 308(a) (1996).
[43] Restatement Second, Agency § 329 (1958).
[44] For the same point, focused on English labor and company law, see Paul L Davies, *Introduction to Company Law* (New York, Oxford University Press, 2002) 38.

questions raised by statutes that incorporate common-law elements, as discussed above, or by common-law decision making that turns to statute as a source of principle. However, by implication, arguments that disfavor statutes imply that one should be skeptical about changes in common-law doctrines that are grounded in principles derived from statutes.

Both defenders and critics of statutes and the legislative process may paint with brushes that are too broad, at least for purposes of questions addressed by this paper. Statutes vary widely and in many dimensions. The Uniform Commercial Code, for example, is the product of long years of effort by drafters whose expertise is incontestable, in a context of relatively open debate.[45] Each article of the Code thoroughly explores its subject matter. The Code has been widely adopted by state legislatures. At the other end of the spectrum, a statute may be so poorly drafted that even its literal meaning is in question. It may have been adopted hastily or by legislators in the thrall of special interests. If only one or a few states adopt the statute, it may not reflect generally shared policy judgments. Moreover, deficiencies in how the statute is drafted may reflect the enacting legislature's inability to make a particular policy choice, with the consequence that the statute and its legislative history may not provide useful guidance about the legislature's intention to courts.

Defenders of statutes tend not to focus on more parochial or otherwise problematic instances of legislative activity. Writing in 1908, Roscoe Pound characterised legislation as 'the more direct and accurate expression of the general will,' in contrast with the common law.[46] All of Pound's specific examples of legislation are ones he applauds on their merits, though, such as married women's property acts, mechanics' lien laws, and statutes that abolish common-law disqualifications of witnesses.[47] Writing more recently and in a more philosophical vein, Jeremy Waldron grounds his account of the dignity of legislation in a characterisation of legislation as an achievement to be celebrated, as 'concerted, cooperative, coordinated activity in the circumstances of modern life … when a large population act together in some common concern even though they disagree among themselves what exactly is to be done.'[48] Although this may be so as a general matter, Waldron's account does not emphasise legislation that few would celebrate.[49]

[45] The original Code is treated as a paradigm instance of legislation that courts should not ignore as a source of law in Roger J Traynor, 'Statutes Revolving in Common-Law Orbits', 17 *Cath U L Rev* 424 ('[t]he final draft was of a piece and it had the look of having been out in the open. It was soon apparent that it also travelled well, as one state legislature after another adopted it.').
[46] Roscoe Pound, 'Common Law and Legislation', 21 *Harv L Rev* 383, 406.
[47] See *ibid*, at 405.
[48] Jeremy Waldron, *The Dignity of Legislation* (New York, Cambridge University Press, 1999) 156–7.
[49] The legislative accomplishments that Waldron mentions are '[e]nterprises like protecting the environment, operating a health care system, securing the conditions for the operation of a market economy, or providing a basis for dispute resolution [which] will founder unless people act in concert, following rules, participating in practices, and establishing institutions.'). *Ibid*, at 156.

In contrast, other contemporary writers may treat legislation at its most parochial as the prototype instance that dominates how even unproblematic legislation should be viewed.[50] Judge (then-Professor) Frank Easterbrook wrote that '[a]lmost all statutes are compromises, and the cornerstone of many a compromise is the decision, usually unexpressed, to leave certain issues unresolved.'[51] If a statute reflects the result of a bargain among competing interest groups, its impact should be limited to the terms of the bargain as expressed in the language of the statute. Even when a statute serves public interests more broadly, a court should not expand its reach unless the legislature expressly signals in the language of the statute that it so intends.[52] In this view, it is the compromise that legislation often represents—a quality celebrated by Waldron—that undermines its normative strength.

Moreover, this argument ignores the many instances in which courts have treated statutes as a reputable source of principled articulations of policy. It is beyond question that courts have often used statutes as a basis for developing common-law doctrines. This practice has a long lineage. For example, to determine when an easement should be established on the basis of long usage of land, English courts in the fourteenth century turned by analogy to 1189, the year set by the Statute of Westminster as the limit of time in which a plaintiff in a Writ of Right could trace title.[53] Later courts adopted proof of twenty years of continuous use to help establish such an easement, drawing an analogy to the twenty-year period of limitations set in a statute of James I.[54] Moreover, statutes continue to shape how courts develop common-law doctrines. Characterising the evolution of contract-law doctrine, Hugh Collins notes that English courts did not dismiss regulation of consumer contracts as 'irrelevant to the normative orientation of private law, but rather private law sought to reconfigure itself in order to place this regulation within its discourses as far as possible.'[55]

[50] Many analyses of the legislative process emphasise the rent-seeking behavior of special interest groups, who 'attempt to procure legislation that transfers wealth (economic 'rents') in excess of what members of such groups could earn in the competitive marketplace to themselves from the public at large.' Jonathan R Macey, 'Cynicism and Trust in Politics and Constitutional Theory', 87 *Cornell L Rev* 280, 294 n 50. Some writers treat politicians as relatively passive brokers of such wealth transfers, while others assign politicians a more actively entrepreneurial role. Compare George Stigler (ed), *Chicago Studies in Political Economy* (Chicago, University of Chicago Press, 1988) with Fred S McChesney, *Money for Nothing: Politicians, Rent Extraction, and Political Extortion* (Cambridge, Mass, Harvard University Press, 1997) 7–13. For a useful comparison of these perspectives, see Todd J Zywicki, *The Rise and Fall of Efficiency in the Common-Law: A Supply-Side Analysis*' (unpublished manuscript, available at http://ssrn.com/abstract_id=326740).

[51] Frank H Easterbrook, 'Statutes' Domains', 50 *U Chi L Rev* 533, 540.

[52] *Ibid*, at 542–3 (giving Sherman Antitrust Act, written in broad terms, as instance in which a statute 'effectively authorize[s] courts to create new lines of common law.').

[53] See Traynor, above n 45, at 406.

[54] See *ibid*, at 406–07.

[55] Hugh Collins, *Regulating Contracts* (New York, Oxford University Press, 1999) 49.

Courts in the United States, likewise—or it is argued, to an even greater extent[56]—use statutes as a normative resource for shaping common-law doctrine. For example, although Married Women's Statutes accorded women power to hold and convey property and power to sue and be sued, courts subsequently drew from the statutes a broad principle of distinct personhood and applied it to justify jettisoning common-law doctrines that presupposed unity of personhood between spouses.[57] Alternatively, a court could decline to draw from a statute a broader normative principle with which to reshape common-law doctrine. However, part of the common-law judicial function is conceded to be 'upkeep' of the substance of common-law doctrines.[58] It would be surprising to exclude statutory materials categorically, either as a basis for recognising that the time for upkeep has arrived or as a basis from which relevant content may be drawn.[59] Thus, if legislatures adopt statutes against a background in which courts often draw on statutes in developing the common law, the fact that the legislature does not authorize the practice in the text of a particular statute does not imply disapproval of it.

One striking feature of much of this literature is the prominent role that metaphors play in it. The metaphors themselves vary. When statutory development is characterised as an 'incursion'[60] or as an 'alien element,'[61] the implied metaphor may be a viral or bacterial infection that afflicts or invades the body of the common law. Alternatively, these usages may reflect the territorial metaphor also implied by the search for a statute's proper 'domain.'[62] Characterising statutes and the common law as 'oil and water'[63]

[56] See P S Atiyah, 'Common Law and Statute Law', 48 *Mod L Rev* 1, 27.
[57] See Traynor, above n 45, at 413–14. See *People v Pierce*, 395 p 2d 893 (Cal 1964) (holding that spouses are capable of conspiracy, overruling *People v Miller*, 22 p 934 (Cal 1889); '[t]he fictional unity of husband and wife has been substantially vitiated by the overwhelming evidence that one plus one adds up to two, even in twogetherness.'); *Self v Self*, 376 P2d 65 (Cal 1962) (holding that one spouse may recover against the other in tort).
[58] See *People v Pierce*, 395 P2d at 895.
[59] See Traynor, above n 45, at 416 (arguing that '[i]t would be wasteful for courts not to utilize such statutory materials when they are readily available for analogy….'). Justice Traynor's argument encompasses at least some forms of 'adoption' by a court of legislation that has not yet become effective. See *ibid*, discussing *Clinkscales v Carver*, 136 P2d 777 (Cal 1943) (although penal statute governing consequences of posting stop-sign at intersection was inapplicable to defendant due to defect in publication of statute, in civil action court treats standard established by statute as appropriate test of conduct to determine civil liability).
[60] See, eg, *Weintraub v Weintraub*, 395 So 2d 302, 303 (Fla App 1981) (antenuptial agreement does not constitute a bar to alimony claim when one prospective spouse did not disclose extent of his resources to other prospective spouse; court limits statute validating antenuptial agreements despite nondisclosure to probate proceedings, given placement of statute within probate code, characterizing it as a 'statutory intrusion upon the common law.').
[61] See Pound, above n 46, at 385 (characterising 'narrow and illiberal view of legislation' as one that 'regard[s] it as out of place in the legal system, as an alien element to be held down to the strictest limits and not to be applied beyond the requirements of its express language.').
[62] See Easterbrook, above n 51.
[63] See Jack Beatson, 'The Role of Statute in the Development of Common Law Doctrine', 117 *LQR* 247.

reflects a chemical metaphor, while physics and astronomy may explain the metaphor implied by discussion of a statute's 'gravitational force'[64] and by contrasting the 'skymark' of the UCC with the 'less spectacular planets' that many statutes represent.[65] Biological metaphors abound as well—the legal system has been characterised as an 'organic whole'[66] encompassing both statutory and common law, while some denigrate the normative significance of statute as a source of law because it 'lacks roots.'[67]

And what is to be made of the heavy use of metaphors? Perhaps only that figurative language may distract the reader's attention from the assumptions that an author makes. In particular, metaphors of invasion and territoriality assume that statutes and the common law are in fact isolated, if not antagonistic, elements in a legal system. Invasion metaphors, in particular, tend to prejudge the substantive question through connotations that suggest disapproval of the invader and support for the invadee.[68] In many instances, however, patterns of mutual reference, influence, and exchange are evident. As discussed at length in section II, statutes often use terminology and concepts drawn from the common law, and courts may apply common-law doctrines to complete interstices within statutory schemes. The common law thus contributes an underlying coherence of meaning. Correspondingly, statutes may influence how courts develop common-law doctrines.

The most apt metaphor may be one of hybridization—the repeated exchange of ideas and approaches between two populations. Hybridisation is not, however, the same phenomenon as convergence, in which two become one, because elements of systems with hybridised features may remain distinctive.[69] Hybridised exchange among legal systems—or between components of a single legal system—can reflect a 'complex web of borrowings of particular features applied to different problems,'[70] a dynamic process that may not be tidy. Evaluating the consequences of hybridised borrowing requires examining instances individually on their merits, as opposed to embracing or condemning them on the basis of their lineage. Another potential pitfall of hybridised exchange is illustrated by

[64] *Ibid*, at 250.
[65] See Traynor, above n 45, at 422.
[66] See Pound, above n 46, at 386.
[67] See *ibid*, at 404–05.
[68] For a thoughtful discussion of metaphorical expression in legal discourse and the functions it may serve, see Michael Boudin, 'Antitrust Doctrine and the Sway of Metaphor', 75 *Geo LJ* 395. In particular, the 'bottleneck' metaphor in monopolisation cases shapes the substantive doctrine because '[t]he immediate subjective associations of 'bottleneck' are almost all unfavorable, for the term summons up pictures of the overlong checkout line at the supermarket or the access ramp to the bridge clogged with cars.' *Ibid*, at 403.
[69] The phenomenon of hybridized legal exchange is explored in the context of comparative environmental regulation in Jonathan B Wiener, 'Whose Precaution After All? A Comment on the Comparison and Evolution of Risk Regulatory Systems', 13 *Duke J Comp & Int'l L (special Issue 2003).*
[70] See *ibid*, at 234.

the 'legal alphabet' metaphor, discussed in the prior section; hybridised exchange may not lead to satisfactory results if it occurs through a process that discourages careful inquiry.

IV. CANDIDATES FOR FURTHER DEVELOPMENT IN COMMON-LAW AGENCY

While some doctrines within the common law of agency are prime candidates for reform on the basis of principles to be drawn from statutes, others are not. Candidates for reform fall into three distinctive categories: (1) instances in which a change in common-law doctrine can be justified on the basis of a statute that is expressly applicable to a closely analogous situation; (2) instances in which a common-law doctrine should be updated to reflect a shift in underlying policy judgments as manifested in legislation; and (3) instances in which non-uniform legislative change has undermined a common-law doctrine as a basis on which parties may form stable expectations. There may be other candidates as well, of course, that lie beyond the scope of this chapter. This section of the paper examines one example in each of these categories. It concludes with an example that illustrates a limiting case, in which it would not be justifiable to draw on a statute as a basis to alter a particular common-law doctrine.

A. Analogous Situations

Consider first an example of a common-law doctrine that warrants change on the basis of a statute that is expressly applicable to an analogous situation. As noted above, within common-law agency, an agent may act with authority on behalf of an undisclosed principal in dealing with third parties. In particular, when an agent enters into a contract with a third party on behalf of an undisclosed principal, the principal and the third parties become parties to the contract, as does the agent. This result is not consistent with the expectations of the third party, who, unaware that the agent represents anyone other than the agent's own self, expects that only the agent will become a party to the contract.[71] A basic consequence of the doctrine is that the principal may require the third party to render performance to the principal and not the agent by making the principal's existence known and by making such a demand. The third party is not obliged to render performance to the principal when the contract made with the agent requires

[71] Indeed, the agent does become a party to the contract. Although the principal's liability on the contract exceeds the third party's expectations, the third party's liability to the principal is beyond them as well.

the third party to perform personal services or rendering performance to the principal would materially increase the burdens borne by the third party.

Early and influential authority for this basic consequence is *Scrimshire v Alderton*.[72] *Scrimshire* continues to represent a prototypical situation, in which a heretofore undisclosed principal—an oat farmer in *Scrimshire*—demands payment from the purchaser of goods that have been sold by an agent on behalf of the principal. Following the principal's demand for payment, the purchaser in *Scrimshire* paid the agent. The agent did not remit the payment to the principal. This sequence of events did not satisfy the purchaser's obligation to pay the principal, now disclosed, with the consequence that the third party paid twice for the oats—once to the agent and once again to the principal following the court's judgment.

In situations comparable to *Scrimshire*, an undisclosed principal's revelation of its existence may confront a third party with a difficult dilemma. Suppose the third party doubts whether it is true that the person demanding payment was the principal on whose behalf the agent acted. If the third party pays the person who demands payment notwithstanding doubts about the legitimacy of the claim for payment, the third party's payment will not discharge the third party's payment obligation if the doubts turn out to be justified. The third party will still be obligated to pay the 'agent,' if that person acted on behalf of no principal, or to pay the person who in fact was the agent's undisclosed principal. Indeed, the third party may pay both, if the third party pays the agent following demand for payment from yet another person claiming to be the agent's undisclosed principal whose claim is legitimate.

On the other hand, if the third party refuses to pay the now-disclosed principal, *Scrimshire* establishes that payment to the agent does not discharge the third party's obligation to the principal. Moreover, if the third party errs in deciding whom to pay, the consequence will be a default on an obligation owed to someone, which may carry collateral consequences under the contract, such as the imposition of late-payment charges, or under other contracts that contain cross-default provisions.

The third party's dilemma is comparable to that of an account debtor whose account is assigned who may well, when the assignee demands payment of the account, doubt the validity of the claim. Revised UCC § 9–406(c) addresses the account debtor's dilemma by providing that the account debtor may request 'reasonable proof that the assignment has been made,' which the assignee must seasonably provide. Unless the assignee does so, the account debtor may discharge its obligation by paying the assignor.[73] A third party who deals with an agent who represents an undisclosed principal is not situated identically to an account debtor whose

[72] *Scrimshire v Alderton*, 2 Strange 182, 93 Eng Rep 1114 (KB 1743).

[73] Revised UCC § 9–406(c). To substantially the same effect is § 9–318(3) in original Art 9.

account is assigned because the agent's dealings, if authorised, immediately create rights and obligations in the undisclosed principal on whose behalf the agent acted. However, this difference is irrelevant to the third party's dilemma when confronted with a demand for payment from a person claiming to be a now-disclosed principal.

Common-law agency doctrine should accord comparable treatment to resolve the third party's dilemma. Section 9-406(c) reflects a principle broader than the assignment context, which is to provide a mechanism for resolving doubts prior to payment. Moreover, deriving and applying such a broader principle does not run counter to fundamental doctrines in agency. In particular, a mechanism comparable to that created by § 9-406(c) would not require a would-be undisclosed principal to sacrifice the benefit it may perceive in dealing through an agent on an undisclosed basis because the mechanism would become operative only *after* an undisclosed principal reveals its existence. The mechanism requires only that a heretofore undisclosed principal provide reasonable proof upon demand of the legitimacy of its claim for payment, a requirement of confirmation that is inapplicable unless and until the principal reveals its existence and identity by demanding payment from the third party.

B. Shift in Underlying Policy Judgment

Statutes may also reflect a shift in societal judgments about how a particular risk should be allocated, in contrast with a doctrine in common-law agency. Consider in this connection the consequences ascribed by common-law agency to a principal's death or loss of capacity. The common-law doctrine terminates an agent's actual and apparent authority upon the principal's death or loss of capacity, although neither the agent not the third party with whom the agent dealt had notice of the principal's death or loss of capacity.[74]

This doctrine is ripe for reconsideration. Of course, an agent's ability to bind a principal's estate through a postmortem transaction is subject to law applicable to the decedent's estate, and an agent's authority to act on behalf of a principal who lacks capacity will be subject to law applicable to any guardian or conservator appointed for the principal. These reservations aside, consider the potential consequences for the agent created by the common-law doctrine concerning the effect of the principal's death. Unaware that the principal has died, an agent may in good faith continue to interact with third parties, purporting to represent the principal. The agent risks claims by third parties that the agent breached the agent's implied warranty

[74] See Restatement Second, Agency § 120 (principal's death automatically terminates agent's actual and apparent authority) and § 122(1) (consequences of principal's loss of capacity equated to consequences of principal's death).

of authority, plus any express warranty of authority to bind the principal that the agent may have given. An agent is liable—usually for damages measured by the third party's expectations[75]—when the principal whom the agent purported to represent is not bound by the agent's action. To mitigate against this risk, an agent might seek confirmation that the principal is alive before consummating a given transaction, but multiple confirmations could be required if the relationship entails either multiple transactions or a single multi-step transaction. Moreover, an agent cannot eliminate the risk entirely through confirmations obtained from the principal because the principal's death following the last confirmation will automatically terminate the agent's authority. At some point, endlessly seeking and receiving confirmations from the principal could become costly and annoying.[76] Likewise, a third party with whom an agent deals on behalf of a principal may believe in good faith that the agent acts with authority, even though the principal has died unbeknownst to the third party (or the agent). Seeking continual confirmations from the principal of the principal's longevity could be costly for all three parties.

The common-law doctrine concerning the effect to be given to the principal's loss of capacity also carries troubling implications. The doctrine may be counter to the principal's intention while the principal has capacity. Prior to losing capacity, a principal may wish that an agent will have authority to act on the principal's behalf when the principal no longer has legal capacity. Moreover, the common-law doctrine places the agent and the third party in the position of assessing the principal's capacity, when no court has determined that the principal lacks capacity, and the principal may believe him or herself to be fully competent. Loss of legal capacity, unlike death, is not always final and its occurrence is not always precisely associated with a particular moment or event that can be understood by lay people who are not experts in law or medicine.

The implications of this aspect of the common-law doctrine were most sharply illustrated by cases testing whether a decedent was the owner of a flower bond at the time of death. Issued by the US Treasury between 1918 and 1971, flower bonds had a long maturity and a relatively low coupon rate; they 'blossomed' or 'bloomed' into their full principal amount, plus any accrued interest, for the purpose of paying federal estate taxes due on the owner's estate. As structured, the bonds created an option in the owner's

[75] See *ibid*, § 329, Comment *j*. Accord, Reynolds, above n 13, para 9–060. A few cases limit the third party to a tort measure of damages, specifically excluding recovery for lost expectations. See, eg, *Griswold v Haas*, 210 SW 356 (Mo 1919); *Martha A Gottfried*, *Inc v Amster*, 511 So 2d 595, 598 (Fla App 1987).

[76] On the potential cost of rules that require multiple confirmations of an agent's authority, see *Kidd v Thomas A Edison*, *Inc*, 239 Fed 405, 408 (SDNY), aff'd, 242 Fed 923 (2nd Cir 1917) ('The very purpose of delegated authority is to avoid constant recourse by third persons to the principal, which would be a corollary of denying the agent any latitude beyond his exact instructions.').

estate to put the bond to the Treasury, and a duty in the Treasury to redeem the bond for the purpose of paying estate taxes, if the owner's estate exercised the option.[77] The unusual configuration of investment characteristics bundled together in a flower bond made it an attractive investment for relatively wealthy taxpayers for whom the prospect of paying estate taxes was an imminent as opposed to a remote concern. As a consequence, it was likely that flower bonds might be purchased on behalf of taxpayers whose capacity as principals was not beyond question. In several cases, the Internal Revenue Service argued that the decedent on whose behalf an agent bought a flower bond was past the point of capacity at the time of the purchase, although no adjudication so established. Courts rejected the argument on the basis that any lack of capacity on the part of the principal would make the purchase voidable, not void.[78] From the standpoint of the deceased principal's estate, this was a desirable outcome because it preserved the tax benefits promised by ownership of a flower bond. However, from the standpoint of risks borne by agents, the resolution is less than optimal. The agent continues to bear the risk that the principal's estate will, for whatever reason, seek to avoid the purchase, or that a still-alive principal will disaffirm the transaction.[79]

The common-law doctrine that creates these issues is not consistent with the policy judgments reflected in many contemporary statutes, including the UCC, the Uniform Durable Power of Attorney Act, and contemporary partnership legislation. All reflect a policy of protecting agents and third parties who deal in good faith without notice of a principal's death or loss of capacity. Under UCC § 4–405, a bank continues to have actual authority to pay or otherwise handle items on behalf of a customer, notwithstanding the customer's death or loss of capacity. The bank's authority ends when bank has knowledge of the customer's death or of an adjudication that the customer has lost capacity.

The Uniform Durable Power of Attorney Act permits a principal to create a durable power of attorney by executing an instrument that appoints an agent whose authority expressly will survive the principal's loss of capacity or become effective upon loss of capacity.[80] Moreover, the Act also provides that when a principal has executed *any* power of attorney, whether or not

[77] See *Weld v United States*, 55 F3d 623, 625 (Fed Cir 1995).

[78] See *Campbell v United States*, 657 F2d 1174 (Ct Cl 1981); *United States v Manny*, 645 F2d 163 (2nd Cir 1981).

[79] In contrast, English authority has long rejected the proposition that a principal, upon regaining capacity, may disaffirm transactions made by an agent on the principal's behalf while the principal lacked capacity. See *Drew v Nunn*, [1879] 4 QB 661 (Ct App) (third-party tradesman who supplied goods, on order of principal's wife who was held out by principal as agent with power to deal and draw checks on his behalf, was unaware of principal's insanity; upon regaining sanity, principal may not disaffirm transactions made by wife-agent).

[80] Unif Durable Power of Attorney Act § 5–502. All states have enacted some form of durable power of attorney statute, whether or not based on the Uniform Act. See Carolyn L Dessin, 'Acting as Agent Under a Financial Durable Power of Attorney: An Unscripted Role', 75 *Neb L Rev* 574, 579.

durable, the principal's death or loss of capacity does not revoke the agent's authority as to the agent and as to third parties who deal with the agent under the power in good faith.[81] Over half of the states have legislation consistent with the effect that the Act prescribes for the effect of a principal's execution of any power.

Contemporary partnership statutes, additionally, provide that a partner's death or loss of capacity shall not automatically dissolve the partnership. Under Revised Uniform Partnership Act § 601(7) (i), a partner's death dissociates the partner from the partnership, as does the appointment of a guardian or conservator under § 601(7) (iii). Dissolution of the partnership follows only if at least half of the other partners decide to wind up partnership business.[82] Moreover, a partnership is bound by a partner's act following dissolution, even when the act is not incident to winding-up, if the act would have bound the partnership prior to dissolution and the other party to the transaction did not have notice of the dissolution.[83]

The policy choice reflected in these statutes is to mitigate the risk otherwise imposed by common-law doctrine on agents and third parties who deal in a manner consistent with a principal's manifestations of assent to be bound by the agent given prior to the principal's death or loss of capacity. These statutes protect agents and third parties who deal in good faith and in conformity with the principal's prior instructions. In contrast, the common-law doctrine focuses on the personal relationship that agency creates between an agent and a principal, terminating the relationship for all purposes when the principal no longer exists or no longer may exercise effective control over the agent. The statutes reflect the judgments that the law need not be so categorical, and that nothing of value is lost by moderating the consequences of death or loss of capacity in some circumstances, in particular to protect reasonable expectations of parties, including agents, who have dealt in a manner consistent with a principal's prior manifestations of assent to be bound by actions taken by an agent. These judgments are strong candidates for translation into common-law doctrine.[84]

[81] Unif Durable Power of Attorney Act § 5–504.

[82] Rev Unif Partnership Act § 801(2)(i).

[83] *Ibid*, § 804(2). In contrast, under the original Uniform Partnership Act (1914), a partner's death automatically dissolved the partnership, thereby terminating all partners' authority except to complete transactions already begun and to wind-up partnership business. See Unif Partnership Act §§ 31(4) and 33. Even the 1914 partnership statute, however, did not parallel the common law doctrine on the effect of loss of capacity. Under Unif Partnership Act § 31(1) (a), the fact that a partner 'has been declared a lunatic in any judicial proceeding or is shown to be of unsound mind' is a basis on which a court may decree dissolution on a petition by or for a partner. Unlike the common-law doctrine, judicial involvement was requisite.

[84] See Restatement Third, Agency §§ 3.07(2) & 3.08(1) (Tentative Draft No 2, 2001). The leading case, apart from the agency context, for the proposition that widespread legislative change may warrant change in common-law doctrine is *Moragne* v *States Marine Lines*, *Inc*, 398 US 375, 390–2 (1970) (holding that in general maritime law, cause of action exists for wrongful death caused by violation of maritime duty; widespread statutory rejection of refusal

C. Statutes that Undermine Bases for Stable Expectations

Statutory developments, even when far from uniform, may undermine the likelihood that a common-law rule will lead parties to form stable expectations around which transactions will be structured. Thus undermined, the common-law rule itself may not survive. Such may become the fate in the United States of the doctrine known (to we happy few aware of it) as the 'sealed contract rule.' Under the sealed contract rule, an undisclosed principal is not liable as a party to a sealed instrument and does not acquire rights under such an instrument.[85] As a consequence, if a contract is made under seal, only the parties identified in the written sealed instrument acquire rights and obligations under the contract. A close corollary of the 'sealed contract' rule, applicable to negotiable instruments, excludes an undisclosed principal as a party who could be liable on a negotiable instrument.[86] The rule acknowledges, of course, that a statute could provide otherwise.[87]

The sealed contract rule was recently affirmed by the Canadian Supreme Court in *Friedmann Equity Developments Inc v Final Note, Ltd* ('*Final Note*') in an opinion that acknowledges that the rule may be at odds with observed decline in the significance accorded to the practice of sealing.[88] However, *Final Note* ascribes a function to the rule, which is to provide a mechanism to exclude undisclosed principals as parties to contracts.[89] This portion of the court's opinion does not explain why this function of the rule would not be served as well (or better) by an explicit provision in a contract that excludes an undisclosed principal as a party, a possibility mentioned earlier in the court's opinion.[90] Additionally, the court characterised the rule as part of a 'system of rules which relate to sealed instruments' and stated that to abolish it 'would thus amount to a fundamental reform of the common law rather than an incremental change.'[91] Change in a

to allow recovery for wrongful death is a 'legislative establishment of policy [that] carries significance beyond the particular scope of each of the statutes involved. The policy thus established has become itself a part of our law, to be given its appropriate weight not only in matters of statutory construction but also in those of decisional law.').

[85] See Restatement Second Agency §§ 191 and 296.

[86] See *ibid*, § 152 (undisclosed or partially disclosed (unidentified) principal not liable as a party to a negotiable instrument that did not name the principal).

[87] See *ibid*, §§ 191 and 296.

[88] See [2000] 188 DLR (4th) 269, 292. Indeed, one commentator on *Final Note* observed that '[i]t will come as a surprise to many lawyers that there exists a common law rule known as the 'sealed contract rule.' Its existence is obscure and its rationale is even more so: Geoff R Hall, 'Preserving the Clavicle in the Cat: Stunted Reform of Common Law Rules in the Supreme Court of Canada', 36 *Can Bus LJ* 89.

[89] See *ibid*, at 293.

[90] See *ibid*, at 279. For further discussion of *Final Note* and this point, see Anthony J Duggan & Kent Roach, 'A Further Note on Final Note: The Scope and Limits of Judicial Law Making', 36 *Can Bus LJ* 115, 121.

[91] [2000] 188 DLR (4th) at 293–94. A subsequent provincial case neither cited *Final Note* nor followed its cautious approach to change in long-standing common-law doctrine. In *Taylor v Scurry-Rainbow Oil* (Sask) Ltd, [2001] 203 DLR (4th) 38 (Sask CA), the issue was whether

common-law rule would be warranted, in the court's view, '[w]hen the hardship which a rule causes becomes so acute and widespread that it outweighs any purpose that it may once have served … there must be evidence of a change in commercial reality which makes such a change in the common law necessary.'[92] The consequence in *Final Note* left the plaintiff, a mortgagee, without an effective remedy to enforce its rights under the contract following default in payments due the mortgagee; the mortgage was made under the corporate seal of an agent on behalf of its undisclosed principals, against whom the sealed contract rule endorsed by the court precluded recovery.[93]

In the United States, the sealed contract rule was last the focus of debate in the 1920's. In *Crowley v Lewis*, the New York Court of Appeals reaffirmed the rule in 1924.[94] Although not the author of the opinion in *Crowley*, Benjamin Cardozo defended the case in a public lecture on the basis that the sealed contract doctrine had been applied so recently in other cases that to depart from it in *Crowley* would be 'difficult if stare decisis was not to be abandoned altogether.'[95] Moreover, Justice Cardozo argued, in *Crowley* '[t]here were other and deeper grounds of policy. Men had taken title in the name of 'dummies,' and through them executed deeds and mortgages with the understanding, shared by the covenantees, that liability on the covenant would be confined to the apparent principal.'[96] Any retroactive change in the rule would thus be unjust; any change should come through legislation.[97] As it happens, New York's legislature abolished the rule a few years later.[98]

Other statutory developments in the United States since 1924 call into question whether the sealed contract doctrine is a basis on which reasonable parties form stable expectations about the significance of providing that a contract is made under seal. In many (but not all) states, statutes abolish any

the Rule Against Perpetuities (RAP) invalidated an oil and gas top lease. A majority of the court held that the RAP did not invalidate the lease, stating that '[s]ince common law rules are judge-made rules, the Court can make exceptions when changing conditions so mandate.' *Ibid*, at paras 96–76. The opinion does not detail the scope of the exception that it envisages for the RAP. For further discussion and critique of the case, see Geoff R Hall, 'Was *Final Note* Not the Final Word? *Scurry-Rainbow* and the Continuing Quest for Balance in the Reform of Common-Law Rules', 37 *Can Bus LJ* 229.

[92] *Ibid*, para 46.
[93] One commentator questions 'whether justice was served by this decision.' See GHL Fridman, 'Some Words About Deeds', 81 *Can Bar Rev* 69, 95.
[94] See *Crowley v Lewis*, 239 NY 264 (1924).
[95] See Benjamin N Cardozo, *Cardozo's Paradoxes of Legal Science* (New York, Columbia University Press, 1928) 70.
[96] *Ibid*.
[97] *Ibid*, at 70–1.
[98] See NY Gen Constr Law § 44-A (except when statute provides otherwise, presence or absence of seal on written instrument without legal effect for instruments executed after 31 August 1941).

general distinction between sealed and unsealed contracts.[99] In a substantial minority of states, statutes eliminate the need for seals on conveyances of property.[100] Additionally, UCC § 2–203 makes law respecting sealed instruments inapplicable to contracts for the sale of goods or to offers to make such contracts. Under Revised UCC Article 3, an undisclosed principal is liable on an instrument when the instrument was signed by an agent or representative and the signature would bind the principal if on a simple contract.[101] Thus, although the pattern is not uniform across all states, statutes make the presence or absence of a seal irrelevant for many significant transactions.

Moreover, unlike the *Crowley* and *Final Note* courts, other courts in the United States did not defer to the prospect of legislative change. At least two courts abolished the sealed contract rule in the absence of legislation abolishing any general distinction between sealed and unsealed contracts.[102]

As a consequence, although parties might well be aware of statutes that affirmatively require seals for particular types of transactions, it is much less likely that a common-law rule ascribing a general consequence to the presence of a seal will shape parties' expectations. Many parties may believe that sealing no longer has general operative consequences because statutes like the UCC permit so many substantial transactions to be undertaken without a seal. Parties who are concerned about the position of an undisclosed principal under a contract are also likely to be aware that the question may be addressed explicitly in the contract. Thus, eliminating the sealed contract rule from the common law canon of agency doctrine is not likely to upset settled expectations. Widespread (although not uniform) statutory development has destabilized the common-law rule as an assumed basis on which parties enter into transactions.

D. A Limiting Case

In the three preceding instances, the argument for change in a common-law doctrine is grounded in statutes from which it is possible to extract a principle of general applicability and normative force. As a contrasting case, consider the common-law doctrine that an agent who enters into a contract on behalf of an undisclosed or unidentified principal becomes a party to the

[99] See Restatement Second, Contracts § 94, Introductory Note (1981).
[100] See Eric Mills Holmes, 'Stature and Status of a Promise Under Seal as Legal Formality', 29 *Willamette L Rev* 617.
[101] See UCC § 3–401 (person is not liable on an instrument unless person signed it or 'the person is represented by an agent or representative who signed the instrument and the signature is binding on the represented person under Section 3–402') and § 3–402 (a person is bound by a representative's signature 'to the same extent the person would be bound if the signature were on a simple contract' and regardless of whether the representative signs in the representative's own name or that of the represented person').
[102] See *Nalbandian v Hanson Restaurant & Lounge*, 338 NE2d 335 (Mass 1975); *Harris v McKay*, 122 SE 137 (Va 1924).

contract and liable upon it to the third party. A principal is undisclosed if the third party with whom the agent deals does not know or have reason to know that the agent represents a principal. A principal is unidentified if the third party does not know or have reason to know the identity of the principal represented by the agent.[103]

What effect should be given to the fact that a principal has filed the documents requisite to becoming a limited-liability entity, or that a principal has registered an assumed name under which it does business within a state? Should such official filings, by themselves, be deemed to give a third party 'reason to know' that it deals with an agent for a disclosed principal and that principal's identity? If so, the agent will not be subject to personal liability on the contract even though the agent has not disclosed the identity of the agent's principal to the third party and the transaction between the agent and the third party has occurred in an vacuum devoid of information from which a reasonable person might discern the principal's identity.

If such official filings are deemed to give notice of the principal's identity to the third party with whom an agent deals, the self-protective incentive provided agents by the common law doctrine to disclose the principal's identity to the third party will be sacrificed. As best can be determined, only one court to date has treated such a filing as effective to give notice to a third party. In *Metro Bulletins Corp v Soboleski*, a majority of the court held that filing an assumed or trade name with the state gave notice to third parties that the person with whom they dealt acted only as an agent for a disclosed principal.[104] The majority reasoned that the third party's theory of liability against the agent meant that 'a business being conducted under a trade name could not use that trade name without providing an affirmative disclosure of the entity behind the business in every transaction with the public.'[105] The dissent read the assumed name statute more narrowly, pointing out that '[i]t does not seem to be too much to ask of an individual acting as an agent in a given transaction who wants to avoid personal liability on a particular contract specifically to disclose his principal to the other party.'[106]

Other cases, like the dissent in *Metro Bulletins*, decline to read registration statutes as bases for charging third parties with notice of the identity of an agent's principal.[107] It is relevant that these statutes themselves do not state that third parties who deal with agents of registered principals are charged

[103] See Restatement Third, Agency § 1.04(2) (defining terms).

[104] 620 A2d 1314 (Conn App 1993), certification granted in part, 625 A2d 823 (Conn 1993). The absence of any subsequent history suggests that the parties reached a settlement after the Connecticut Supreme Court agreed to consider this question.

[105] 620 A2d at 1318.

[106] *Ibid*, at 1320.

[107] See, eg, on statutes requiring registration of an assumed name, *J & J Builders Supply v Caffin*, 56 Cal Rptr 365, 369 (App Ct 1967); *Robinson & St. John Advertising & Public Relations*, *Inc v Lane*, 557 So 2d 908 (Fla App 1990). On registration as a limited liability entity, see *Waste*, *Water & Land*, *Inc v Lanham*, 955 P2d 997 (Colo 1998).

with notice of the principal's identity.[108] Instead, the statutes provide a source to which a third party may turn to obtain information about the status or identity of parties with which it deals. The statutes, in other words, do not embody or express a general principle that charges third parties with notice of the identity of an agent's principal when the principal has complied with the statute. Thus, registration statutes do not provide a basis for altering the common-law doctrine defining when a principal's identity has been disclosed.

V. CONCLUSION

The relationship between statutes and common-law doctrine has long been one of mutual exchange and borrowing, not one of necessary antagonism and incompatibility. A recent instance is the incorporation into Louisiana's Civil Code of the doctrine of the undisclosed principal, a hallmark of common-law agency.[109] Common-law terminology and concepts are often a source of coherent meaning across diverse legislative schemes, while statutes may provide a source of general normative principle upon which courts may draw in developing common-law doctrines. It is unsurprising that federal courts turned to common-law agency to define 'employee' in applying federal statutes in which that status triggers operative consequences.[110] However, transplanting common-law doctrines into statutory settings requires care and close attention to statutory purpose and the operative meaning of common-law doctrines. In particular, the metaphor of the 'legal alphabet' of common-law terms has the potential to distract attention from substantive issues that underlie the reasons why transplanting common-law doctrine into a statutory context might or might not make sense.

Courts have long viewed statutes as a source of normative principle that may be useful in developing the common law. Statutes may articulate broad normative principles readily applicable to analogous situations governed by common-law doctrines or may reflect shifts in societal judgements from which common-law doctrine should not be isolated. Statutes may also undermine the likelihood that a common law doctrine—like the sealed contract doctrine—will provide a basis on which parties to transactions develop settled expectations.[111] The contemporary vitality of common-law agency would be enhanced by careful consideration of statutes as bases for further development.

[108] See 955 P2d at 1002 (provision in statute that '[t]he fact that the articles of organization are on file with the office of the Secretary of State is notice that the limited liability company is a limited liability company and is notice of all the other facts set forth therein which are required or expressly permitted to be set forth in the articles of organization' only insulates the company's members and managers from liability based simply on their status as such).
[109] See text accompanying ns 11–13 above.
[110] See text accompanying ns 13, 20–1, & 30–7 above.
[111] See text accompanying ns 85–102 above.

4

Property, Private Government and the Myth of Deregulation

PADDY IRELAND

ONE OF THE defining features of recent years has been the transformation wrought in economic policy and management worldwide by resurgent neoliberalism. The last two decades have seen the ambit and power of 'the market' extended and extended again as market principles and disciplines have been introduced into areas previously untouched by them and intensified in others where they were already present. Financial markets have been transformed, with dramatic consequences, and organisational forms within both the private and public sectors have been reshaped. Outsourcing and subcontracting have become the order of the day and markets and quasi-markets have appeared in sectors in which the operation of market forces had hitherto been limited, most notably perhaps in health, education and the former nationalised industries. These developments have in turn been associated with the erosion of welfare states and significant shifts in the distribution of wealth at both the national and international levels. In recent years the rich have got richer and the poor poorer, and the gulf continues to widen: the wealthiest 1 per cent of the world's population now account for a larger percentage of world income than the bottom 50 per cent.[1]

[1] See Branko Milanovic (World Bank, Development Research Group), 'True World Income Distribution, 1988 and 1993' (2002) 112 *Economic Journal* 51–92 and the United Nations *Human Development Report* (1999). The increases in inequality in the United States, where the income of the richest 1% of the population just about matches that of the poorest 40 per cent, have been so great that the mainstream economist Paul Krugman recently questioned the proposition, taken for granted in America, that because the United States has the highest per capita income among the major countries of the world it follows that ordinary Americans are materially better off than their counterparts elsewhere. In fact, he argues, per capita income is higher in the US only because the rich are so rich. Life expectancy, Krugman observes, is 'well below that in Canada, Japan and every major nation in Western Europe'; indeed, 'male life expectancy is [now] lower in the US than it is in Costa Rica.' One result of this is that the claim that redistributing wealth from rich to poor would make little difference to the material well-being of the less-wealthy is simply wrong. See 'For Richer',

Among the key watchwords of these processes have, of course, been 'privatisation' and 'deregulation', references to the extension of the scope of private property, contract and market mechanisms, and to the free, unhindered operation of those mechanisms once in place. The underlying supposition is that free markets—private economic ordering and the unregulated forces of supply and demand—are the best way to ensure the efficient allocation of resources and to maximise economic growth and wealth for the benefit of society as a whole. Indeed, it has come widely to be believed that failure to privatise and deregulate and to give the free market its head is a recipe for competitive disaster, both at the level of the individual organisation and of the nation state, hence the emergence of the notion of 'competitive deregulation'. These developments have, of course, been accompanied by growing scepticism as to the desirability and effectiveness of intervention by the state in economic (and other) matters—and in some cases even a questioning of its right to do so—with policy-makers increasingly purporting[2] to favour a paring down of state involvement in economic life.[3] Correspondingly, the changes which have occurred are commonly seen in terms of an underlying shift from state to market and from public to private ordering; as allowing people to order their own affairs using contract and the market, with the assistance of their lawyers—if, of course, they have one. One would expect these developments to meet with approval from those branches of the profession undertaking commercial work, if not necessarily from those undertaking work that is (or was) legally aided.

None of this is to say that the proponents of privatisation, deregulation and the market think that there is no place for 'regulation'. Within academia even fundamentalists usually concede that some situations are characterised by 'market failure', and that in these contexts regulation of some sort may be perfectly appropriate. Indeed, this has generated lively debates about the kinds of 'regulatory regimes' which should be adopted in particular contexts, with some drawing policy conclusions which seem markedly interventionist. But more and more of these debates are taking place within a framework in which the superiority of market ordering and market principles is presumed, passing largely unexamined and unchallenged. Indeed, the main purpose of much actual and proposed 'regulation'—even that with a progressive bent—is often not so much to replace market goals

New York Times Magazine, 20 October 2002. Closer to home, a recent report suggested that more than half of the 600,000 children living in inner London are being brought up below the government's official poverty line, see *Guardian*, 19 October 2002.

[2] 'Purporting' because, as I argue below, in crucial respects the regulatory role of the nation state, not to mention that of a range of international agencies, such as the International Monetary Fund and the World Trade Organisation, has never been greater.

[3] See Kerry Rittich, *Recharacterizing Restructuring: Law, Distribution and Gender in Market Reform* (London, Kluwer Law International, 2002) 49–61. This development has, of course, been accompanied in philosophical and legal circles by the rise of the libertarian defences of the minimalist state associated with writers such as Robert Nozick and Jeffrey Paul.

as to assist in their realisation: to perfect the operation of (improperly functioning) markets or to achieve market outcomes ('efficiency') by non-market means. The debates are also increasingly underlain by presumptions about the universality of *homo economicus* and rational egoism. One paradoxical consequence of the lack of trust which results from these assumptions about human motivation is that in areas where market disciplines are thought to be absent or weak, as in many parts of the public sector, vast and expensive state bureaucracies are being created to monitor and audit. *Homo economicus* may, as Donald Mackenzie suggests, be the rather impoverished product of an imagined 'market' rather than of real-world market*s*, but his influence in the shaping of public institutions and the social world more generally is growing apace.[4] One result has been that Universities in the UK have increasingly been caught between a rock and a hard place, subject to both overweening state bureaucracies and growing market pressures. Gradually, their structures and goals have become shaped less by pedagogical considerations than by the need to make money following cut-backs in state funding, by the search for ways to teach growing student numbers on a declining unit of resource, by the need to comply with bureaucratic drives for so-called 'quality assurance', and by the demands for greater research productivity and quality (the RAE). Increasingly, universities have been commercialised and higher education commodified, trends likely to be intensified by the recently announced changes to tuition fees and university finance. The rise of neoliberalism has also impacted intellectually, not least in the growing influence on legal scholarship of the law-and-economics movement.

The proponents of markets and deregulation have, of course, met with resistance, and continue to do so, both within academia and beyond. Indeed, in tracing the current debates about the place and value of free trade and free markets, and the proper role of the state in economic and social affairs, one cannot escape a sense of *deja vu*, for they bear a striking resemblance to those which took place in the nineteenth and early twentieth centuries around the issues of laissez-faire, free trade and free competition. These earlier debates, which were especially fierce in the United States during the *Lochner* era, generated the emergence of what has been called 'the first great law and economics' movement.[5] Paradoxically, however, unlike their latter-day counterparts, the members of the original law and economics movement—institutional economists like Richard T Ely, Thorstein Veblen and John R Commons, and legal realists like Robert Lee Hale—were all deeply sceptical not only of the capacity but of the very notion of the 'free' market. They believed not merely that the state had a vital role to play

[4] See Donald MacKenzie, Book Review, *London Review of Books*, 31 October 2002 at 24.
[5] Herbert Hovenkamp, *Enterprise and American Law, 1836–1937* (Cambridge Mass, Harvard University Press, 1991).

in promoting social welfare but that it simply couldn't help but be deeply immersed in economic affairs. In this respect, the work of Robert Hale is particularly interesting and, indeed, relevant to some of the themes of conference, for Hale not only had close personal ties with a number of the members of the Supreme Court—Stone (his former Dean at Columbia), Brandeis, Cardozo, Douglas, Frankfurter, with all of whom he corresponded—but saw his 'central task' as that of 'educat[ing] lawyers, judges, and economists to the role of positive legal entitlements in shaping what were supposed the 'natural' rights of ownership and the 'natural' laws of distribution' in a free market economy.[6] Hale's work spanned the legal and the economic, and the academic and the practical—moving from abstract, philosophical matters such as the nature of coercion and property rights to rather more everyday matters such as public utility regulation.[7] It continues to offer invaluable insights into the nature of 'free' markets and contemporary

[6] Barbara Fried, *The Progressive Assault on Laissez Faire* (Cambridge Mass, Harvard University Press, 1998) 3. Robert Hale (1884–1969) was trained as both an economist (under Taussig) and a lawyer. He practiced for a Chicago law firm and worked for AT&T before moving into academia and the economics and law departments at Columbia. He wrote numerous law journal articles, gathering his ideas together towards the end of his career in a book entitled *Freedom Through Law* (New York, Columbia University Press, 1952), published shortly after his retirement. His work was revived by the critical legal studies movement in the 1980's, most notably by Duncan Kennedy, and has been the subject of a number of assessments. For further biographical details of Hale and accounts of his work, see Neil Duxbury, 'Robert Hale and the Economy of Legal Force' (1990) 53 *Modern Law Review* 421 at 422–24; Warren Samuels, 'The Economy as a System of Power and its Legal Bases: The Legal Economics of Robert Lee Hale' (1973) 27 *University of Miami Law Review* 261 at 263–64; Pavlos Eleftheriadis, 'Unfreedom in a Laissez-Faire State' (1994) 80 *Archives for Philosophy of Law and Social Philosophy* 168; Duncan Kennedy, 'The Stakes of Law, or Hale and Foucault!' in *Sexy Dressing* (Cambridge Mass, Harvard University Press, 1993) 83.

[7] At the time that Hale was writing the seemingly arcane issue of public utility rate regulation was one of the main battlegrounds upon which was fought out the question of the proper limits of state intervention in, and control of, private property. For Hale, the issue made explicit the role of law in determining the distribution of wealth. One of the focal points of his work was the Supreme Court decision in *Smyth v Ames* (1898) 169 US. 466, which held that whether the rate schedule of a company was 'extortionate' depended on whether the return yielded by those rates offered a 'fair return on the fair value of the property'. The test was laid down to determine whether the rates fixed by a state were too low and thus in violation of the company's constitutional right not to be deprived of property without due process of law. Hale pointed out that the test was illusory as it wasn't possible to ascertain the value of 'the property' (whether this meant the company's assets or shares was never entirely clear) without reference to the rates. For Hale, rate-setting thus opened up a whole series of questions of general importance: 'When the policy of competition is abandoned in favor of regulation of rates and prices, the judicial determination of the constitutionality of any particular rate reduction leads the courts logically into an appraisal of the economic justification of the various advantages derived by individuals from the ownership of income-yielding property', see Robert Hale, 'Economic Theory and the Statesman' in R G Tugwell (ed), *The Trend of Economics* (New York, Knopf, 1924) 189 at 194–5. See also Robert Hale, 'Economics and Law' in WF Ogburn & A Goldenweiser (eds), *The Social Sciences and their Interrelations* (London, Allen & Unwin, 1927) 131 at 132–3; 'Rate Making and the Revision of the Property Concept' (1922) 22 *Columbia Law Review* 209. Hale's work in this area was cited approvingly by Justices Black, Douglas and Murphy in key cases (see Samuels, above n 6 at 264) and contributed to the demise of the *Smyth v Ames* formula in the 1940's.

notions of 'regulation' and 'deregulation', as well as providing a powerful critique of neoliberalism.

This chapter will not seek directly to evaluate the pros and cons of the free market or regulation, either in general or in specific contexts. It will try, rather, to undertake the prior task of critiquing these concepts, suggesting that as traditionally understood both are based on misapprehensions as to the nature of property and property rights and, more specifically, as to the role of law in their constitution and maintenance. As a result, it suggests, both concepts are also in certain important respects premised upon misapprehensions as to the nature of markets themselves. Using Hale's work it will explore the legal foundations of property and markets in contemporary capitalist societies, arguing that the private property rights which are the pre-conditions of market trading are legal constructs—the positive ('public') creations of state and law—as indeed are many of the bargaining rules which shape trading activity.[8] Focusing in particular on intangible property, the paper argues, therefore, that both private property and markets are in certain crucial respects the *products* rather than the subjects of legal 'regulation'; and that, as a result, legal regulation is both ubiquitous and inevitable. This, it will suggest, renders highly problematic the public-private, state-market dichotomies and the notions of 'natural', market-based laws of distribution which underlie much contemporary social, economic and legal analyses. Distribution of the social product is not the largely inevitable result of essentially natural economic forces beyond human control, but a phenomenon shaped by law.[9] The chapter thus seeks to remind us of the pervasiveness of fundamental policy choice in law and of the policy choices which are already embodied in it, seeking to expose and to question things which are commonly taken for granted. In so doing, it tries also to remind us of what is at stake in legal policy formulation and legal decision-making, not least in commercial contexts where property rights are often at issue. Paradoxically perhaps, given the themes of the conference, it also makes a case for the maintenance of critical distance between those working in legal academia and those working in commercial practice.

[8] But not all of them. For an illuminating account of the many extra- and non-legal aspects of trading activity and market construction, see Hugh Collins, *Regulating Contracts* (Oxford, Oxford University Press, 1999) 97–126. Collins stresses the importance of trust and non-legal sanctions in the constitution of markets and the existence, therefore, of 'flourishing markets where no effective state power exists to impose legal obligations'. He concedes, however, that a 'normative system of property rights … is presupposed in all trading activity' and identifies private law as the mechanism for allocating property entitlements. In this context it should be remembered that the role of law in relation to property and property rights is not merely allocative but constitutive, for many of the most important markets in the modern world are markets in which legally constructed, intangible objects of property ('things' composed of bundles of rights) are traded. See text below.

[9] In certain respects, therefore, the paper will confirm and elaborate EP Thompson's idea that law is 'deeply imbricated within the very basis of productive relations', *Whigs and Hunters* (London, Allen Lane, 1975) 260–61. On this see Paddy Ireland, 'History, Critical Legal Studies and the Mysterious Disappearance of Capitalism' (2002) 65 *Modern Law Review* 120.

I. MARKETS AND THE UBIQUITY OF COERCION

The term 'regulation' is often used in a manner which suggests the prior, essentially pre-legal and pre-regulatory existence of something that might be regulated, just as the term 'free market' is commonly used so as to suggest the possibility of a spontaneously arising, largely pre-legal, 'private' order of exchange. The image conjured up is one of a world in which the possessors of essentially natural, pre-political, property rights—rights to 'things', including their own labour—engage in voluntarily exchanges with other property owners and negotiate for themselves further contractual rights. In this pre-regulatory world, property is seen as creating discrete, vertical relationships between individuals and the things they own, and contract as creating voluntary, horizontal relationships between individual property-owners. The resulting arrangements are said not to involve domination or coercion, being the allegedly consensual products of market exchanges freely and voluntarily entered into; nor to involve interventions by the state beyond its enforcement of property rights and any rights which have been freely bargained for. They are, rather said to be the products of unregulated, consensual, private activity. 'Regulation' is thus portrayed as something which is imposed—by law—from without.

In general terms, two related legitimations are offered for the arrangements and outcomes which result. The first tends to treat individual liberty as an end in itself and is based upon a separation of processes and outcomes, of law and politics, and marginalises scrutiny of consequences. The arrangements and outcomes generated by market processes, it is argued, are the expressions of individual exercises of freedom, and any state intervention in or regulation of them prima facie constitutes an unwarranted intrusion into individual liberty, a politically non-neutral interference into free expression and the natural operations of the market. The second is more overtly consequentialist and utilitarian: as the products of freedom of contract, it is argued, the resulting arrangements and outcomes are *a priori* 'efficient' and serve, therefore, to maximise social wealth. The legitimations, are, of course, linked in that what allegedly makes the arrangements and outcomes arising out of these consensual market exchanges efficient is precisely that they are the supposed expressions of individual choice and volition: even if we don't want to protect the property rights and freedom of contract for their own sakes, we should do so for reasons of efficiency and competitiveness. Both legitimations figured in the defences of laissez-faire mounted in the late nineteenth and early twentieth centuries, and both figure today, though as a result of the rise of law-and-economics consequentialist, efficiency based arguments tend to predominate. Both were also among the targets of the first wave of progressive institutional (law and) economists. Ruthlessly deconstructing the public-private dichotomy, they questioned the depiction of the market as an essentially natural, private, self-sustaining realm of freedom to be sharply contrasted with 'the state' as a realm of public regulation and coercion. Robert Hale was one of the leaders of the assault.

In his analysis of markets, Hale drew an important distinction between what he called 'voluntary' and 'volitional' freedom, the former signifying complete individual autonomy with no constraints or limits on choice, the latter a circumstantially limited exercise of choice.[10] The market freedom of individuals, Hale argued, is almost invariably volitional, nearly all choice constrained choice, because of the 'background constraints' on behaviour. Some of these constraints emanate from the legal rules structuring bargaining conduct. For example, while the bargaining which takes place between employers and employees is shaped in part by such things as the supply and demand for particular types of labour and its marginal productivity, it is also structured by the many legal rules which influence bargaining between the parties.[11] These range from the rules governing the movement (free or otherwise) of capital and labour to those governing the right of workers to organise, from those concerning the conduct of the internal affairs of trade unions and companies to the general rules of contract and tort, from the rules governing the individual contract of employment to those governing the economic torts, the right to strike, secondary picketing, lockouts, minimum wages, and so on.

According to Hale, however, the most important background constraints are those arising out of property rights—rights which are themselves legal constructs. For these give their holders the crucial power to exclude and withhold, enabling them to determine whether or not the use of things by another is lawful and to impose terms on those others as the price of making it so. To take a simple example, a consumer has to surrender money to acquire a loaf of bread, while a baker has to surrender bread to get money. Both have volitional rather than voluntary freedom in the sense that they are limited by the capacity of the other, through their legally enforceable property rights, to withhold that which they own.[12] When we enter contracts, therefore, we are not in a position to exercise voluntary freedom (completely free choice), 'for our capacity to contract is governed by a pre-existing regime of property relations which determines, first, that we must engage in the process of proprietary exchange in order to survive and, secondly, that our ability to engage in or abstain from this process at any

[10] Hale did not use the idea of voluntary freedom as a norm with which to judge real-world situations; he used it simply to highlight the restrictions on freedom which were always present and the resulting ubiquity of volitional (partial rather than complete) freedom. See Robert Hale, 'Coercion and Distribution in a Supposedly Non-Coercive State' (1923) 38 *Political Science Quarterly* 470.

[11] It should not be forgotten, however, that not all of these rules are legal in nature, see Collins, above n 8.

[12] See Hale, above n 10 472–73.

particular time will be governed by the extent of our bargaining power vis-à-vis others'.[13] Hale did not deny that many of the particular rights and duties created by contracts were created at the initiative of private individuals. He argued, rather, that their 'freedom to decline to do so [wa]s nonetheless circumscribed' in that each party 'makes the contract in order to acquire certain legal rights he does not now possess, or to escape certain legal obligations with which he is now burdened.' Were their liberty not restricted by these obligations imposed by the law and enforced in the courts, people 'might never submit [themselves] to the new obligations of the contract. Thus in a sense each party to the contract, by the threat to call on the government to enforce his power over the liberty of the other, *imposes* the terms of the contract on the other'.[14] Hale thus argued that while the particular rights and duties created by contract are usually created volitionally, it does not follow that they are not in an important sense coerced, for in market exchanges the parties exert coercive pressure on one another by virtue of their capacity to withhold or threaten to withhold their property.[15]

Hale's use of the term 'coercion' in this context was not intended to suggest either the use of overwhelming force or deprivation of all choice. Following Holmes, he noted that even the victim of a robbery—your money or your life—had a choice, albeit an unenviable one.[16] Nor was it intended to carry the 'stigma of impropriety'.[17] The 'undoubtedly coercive character of the pressure exerted by the property-owner', he argued, is not in itself grounds for condemnation, nor does it mean that contracts should be set aside.[18] It simply denotes circumstances that narrow available choice: just

[13] Duxbury, above n 6 433.

[14] See Robert Hale, 'Law Making by Unofficial Minorities', (1920) 20 *Columbia Law Review* 451 at 452; 'Bargaining, Duress and Economic Liberty' (1943) 43 *Columbia Law Review* 603 at 606. See also Fried, above n 6 55–57.

[15] See Hale, above n 10 471 *et seq*.

[16] In *Union Pacific Railroad Co v Pub Serv Comm* (1918), 248 US 67 at 70 Holmes had argued that '[i]t always is for the interest of a party under duress to choose the lesser of two evils. But the fact that a choice was made according to interest does not exclude duress. It is the characteristic of duress properly so called'. Hale himself observed that 'even a slave makes a choice. The compulsion which drives him to work operates through his own will power. He makes the 'voluntary' muscular movements which the work calls for, in order to escape some threat …', See 'Bargaining, Duress …', above n 14 606.

[17] On occasions Hale used other terms to try to convey the same meaning—'compulsion', 'pressure', 'force', 'influence', 'duress' and sometimes even 'oppression', see Samuels, above n 6 280–81; Hale, above n 10 471, 476. There is no doubt that many orthodox economists, intent on highlighting the consensual, free nature of markets, found Hale's insistence on their coercive nature highly irritating and provocative. Hale suggested that it was precisely because the notion of coercion did carry such a stigma that 'the coercive character of many innocent acts is frequently denied', see Hale, above n 10 471.

[18] Hale, above n 10 471. Although it did not follow for Hale that 'all contracts be set aside on the ground of duress', it did mean that one had to decide what sorts of coercion were to be deemed permissible and which not, see Samuels, above n 6 282; Robert Hale, *Freedom Through Law* (New York, Columbia University Press, 1952) 124–5.

as the presence of coercion does not necessarily imply a lack of volition, so the presence of volition does not necessarily imply the absence of coercion. In market exchange situations, rather, each side exercises volitional freedom but is also, to varying degrees, constrained and coerced by the other party. Like volitional freedom, therefore, coercion is *ubiquitous*, though it does not usually take the form of direct physical compulsion but of circumstantial background constraints on choice. Property rights are crucial because in the network of coercive pressures and counter pressures which make up markets 'each pressure consist[s] in the last analysis either of the power to lock or unlock the bars which the law erects against the non-owners of each piece of property, or else of the power to withhold or not to withhold labour'.[19]

Hale concluded that all markets are inevitably characterised by a 'structure of mutual coercion', a structure which, as in large part a reflection of 'the particular legal rights with which the law endows [some], and the legal restrictions which it places on others',[20] is subtly constituted by various background legal rules. The structure is *mutually* coercive because coercion is not one-sided, the power of one party nearly always being countered to some degree by the power of the other. In some contexts, such as those involving ordinary commercial contracts, the element of coercion is often 'sufficiently mutual' to 'prevent one side acquiring … any undue advantage over the other'. But in others, where 'the rights and privileges which one party possesses are vastly superior in strategic importance … the other party may in effect be compelled to submit by contract to almost any terms imposed by the stronger party'.[21] In labour contracts, for example, the balance of power usually lies with the employer, for although workers are under no legal duty to work for a particular employer, there is 'an indirect coercion to work in *some one's* employ'. In this context, much depends upon 'how tightly woven is the network of property duties which surround them, on how scarce a variety of labor each one of them possesses in his own person, and on how much opportunity they have to unite their industrial powers to withhold their labor'.[22] Crucially, Hale argued, the structure of mutual coercion is central to the distribution of wealth and income: changes to it inevitably affect the bargaining power of the parties and therefore bargaining outcomes. Moreover, because the structure is largely legally constituted, law and state cannot but help be deeply implicated in the determination of wealth distribution. Like it or not, he insisted, legislators and judges are inevitably engaged in an ongoing process of confirming and/or adjusting

[19] Hale, 'Economics and Law', above n 7 138. Indeed, Hale argued, it is precisely because power is commonly exercised through property rights that the coercive character of the pressure possessed by property owners is disguised, see Hale, above n 10 474.
[20] Hale, 'Bargaining, Duress …', above n 14 625.
[21] Hale, 'Law Making …', above n 14 452.
[22] Hale, 'Economics and Law', above n 7 137–8.

the background legal rules which are central determinants of the distribution of wealth in a market capitalist society.

II. PROPERTY RIGHTS AS LEGAL CONSTRUCTS

The full significance of Hale's analysis only becomes clear, however, when one examines more closely his claim that property and property rights, the principal sources of coercive power, are themselves legal constructs. As Barbara Fried observes, 'Hale's argument was more than Austinian positivism reworked … he insist[ed] not merely that property rights are what the state says they are … [but] that until the state says something, in its official or unofficial guise, property rights are (conceptually) nothing at all'.[23] Drawing on the work of Wesley Hohfeld and his own work on public utility rate regulation and the valuation of shares, which had sensitised him to the growth in (and peculiarities of) intangible property forms, Hale argued that property and property rights are legal constructs in ways even more fundamental than those noted by the positivists.

Until the later nineteenth century, the prevailing conception of property was essentially 'physicalist', meaning that property was generally conceptualised in terms of concrete, tangible, physical objects and property rights seen as relations between persons and things.[24] As the century progressed, however, the legally constructed nature of property rights came into ever sharper relief (for those who cared to look) as a result of the changes which were taking place in the nature of property itself. The spread of capitalist social relations and the credit system during the eighteenth and nineteenth centuries saw the gradual emergence and spread of 'a dazzling variety'[25] of intangible, financial forms of wealth, many of them titles to revenue. Although originally classified as *choses in action* and commonly lacking free transferability, by the nineteenth century many of these income rights had, in the words of Sir William Holdsworth, 'changed their original character' and become 'very much less like merely personal rights of action and very much more like [freely transferable] rights of property'.[26] In short, they were reified. This led to 'a consistent tendency toward generalisation

[23] Fried, above n 6 80–81.
[24] See Kenneth Vandevelde, 'The New Property of the Nineteenth Century: The Development of the Modern Concept of Property' (1980) 29 *Buffalo Law Review* 325.
[25] Morton J Horwitz, *The Transformation of American Law, 1870–1960* (Oxford, Oxford University Press, 1992) 150.
[26] Sir William Holdsworth, *History of English Law* 7th edn, (London, Methuen, 1969) v VII 543. Marx outlined the conditions in which this character-change happened in his analysis of what he called 'fictitious capital', when he outlined the (historical) conditions in which titles to revenue (such as the share) developed a capital value of their own separate from the value of the productive assets from which their value was derived, see Paddy Ireland, Ian Grigg-Spall & Dave Kelly, 'The Conceptual Foundations of Modern Company Law' (1987) 14 *Journal of Law & Society* 149.

and abstraction of the idea of property', as a result of which the old conceptions derived from land were gradually displaced. Property was 'dephysicalised' and defined much more expansively to embrace not merely tangible objects but any 'thing' with exchange value. Crucially, in this process 'property' came to refer not merely to physical objects but to a variety of reified 'bundles of rights', and anything which interfered with the market value of the latter was increasingly deemed to be an interference with property—in the United States a potential 'taking' under the constitution.[27] This expansion of the conceptual boundaries of 'property' enabled the wealthy to obtain greater protection for their rights and interests but it also soon threatened to become the source of property's conceptual undoing.[28] In this respect, Wesley Hohfeld was a key figure.

Hohfeld argued that what we loosely refer to as 'rights' are in fact a number of distinct legal capacities or entitlements, which can be, and are, bundled up in very many different ways. He broke these capacities or entitlements down into four groups—right, power, privilege and immunity—and on this basis famously developed a complex typology of jural correlatives in which each legal capacity of a rights holder is defined by a corresponding non-capacity among non-rights holders. Influenced by the growing detachment of the concept of property from concrete, tangible objects, he then applied this typology to a range of legal relations, including the rights *in rem* traditionally understood as rights to 'things' (and therefore as 'property' rights), arguing that these too were actually complex bundles of rights, privileges, powers and immunities. Properly conceptualised, Hohfeld insisted, rights *in rem* were not rights to things but were, rather, like all rights, rights against persons, any right over a tangible thing entailing a duty owed by someone else to the rights-holder which the state will enforce. Thus the owner of a fee simple in land, Hohfeld argued, holds not a unitary right but a complicated bundle of rights, privileges, powers and immunities: the right to ensure that others ('duty-holders') do not enter the land or cause physical harm to it; a range of privileges, such as to occupy the land or harm it, subject only to any restrictions imposed by public policy; the power to alienate some or all of his legal interests in the land to others; and a number of immunities from any attempts to change his legal capacities in relation to it without his consent. Conceptually, this had potentially very radical implications, for it followed, Hohfeld argued, that rights *in rem* 'establish not

[27] It thus came to be suggested in the US that property was, for the purposes of constitutional protection, 'every valuable interest which [could] be enjoyed as property and recognised as such'. Thus, in the realm of eminent domain, for example, there was pressure to redefine the nature of interference with property rights in highly abstract terms, 'not as an invasion of some physical boundary but as any action that reduced the market value of property', see Horwitz, above n 25 145–147; Vandevelde, above n 24. It was this problem, which reared its head in the cases concerning government regulation of the rates charged by public utilities, that so interested Hale.

[28] See Horwitz, above n 25 at 151–154.

vertical relationships between people and things … [but] a series of horizontal relationships among people, in which each capacity in the owner's entire bundle of rights imposed a correlative incapacity on nonowners'.[29] Property does not, therefore, describe any res or object of sense at all; it is merely a bundle of legal relations.[30] Hohfeld thus suggested that the traditional distinction between rights *in rem* and rights *in personam*, in which the former were seen as involving rights to things and the latter rights over people, had been incorrectly drawn. In his account, the distinction between them was not one of subject-matter but one of scope, of the number of persons affected: rights *in rem* were 'multital', rights *in personam* were 'paucital'. It was the extent of the jural relations that they entailed that distinguished rights as *in rem* rather than their association with some object of ownership. Hohfeld's scheme thus threatened to blur the lines between rights *in rem* and rights *in personam*, and between property rights and contractual rights.[31] All property was composed of legally constructed bundles of rights.

The potential radicalism of this approach was not, however, confined to conceptual matters. By unravelling property into its various, separate, functional relations, Hohfeld made it clear that property and property rights are, at root, *social relations*, a myriad of legally constituted, personal rights between individuals underwritten by the state, something often concealed by their reification.[32] In thereby emphasising that the essence of property rights is their imposition, through law, of correlative (coercive) restraints on nonowners, Hohfeld's scheme prefigured Hale's notion of mutual coercion. It also highlighted the fact that there is no such thing as natural, neutral, pre-political property: property rights, it suggested, are intricate, eminently changeable, legally constructed relations between people.[33] His rather dry analysis thus underlined property's contingent character and further subverted the idea of its specific forms as absolute and natural,[34] a theme

[29] Wesley Hohfeld, 'Some Fundamental Legal Conceptions as Applied in Judicial Reasoning' (1913) 23 *Yale Law Journal* 16; 'Fundamental Legal Conceptions as Applied in Judicial Reasoning' (1917) 26 *Yale Law Journal* 710. See Fried, above n 6 52–53. 'The right of ownership in a manufacturing plant is', Hale observed, 'to use Hohfeld's terms, a privilege to operate the plant, plus a privilege not to operate it, plus a right to keep others from operating it, plus a power to acquire all the rights of ownership in the products', see 'Rate Making ….', above n 7 214.

[30] See Arthur Corbin, 'Taxation of Seats on the Stock Exchange' (1922) 31 *Yale Law Journal* 429.

[31] See the discussion in John R Commons, *Legal Foundations of Capitalism* (New York, Macmillan, 1924) 91 *et seq*.

[32] The potential radicalism of Hohfeld's conception of property is perhaps best highlighted by its similarities with that of Marx who asserted that social relations under capitalism take reified and fetishised forms. Appearances notwithstanding, he argued, capital is not a thing but a social relation.

[33] An insight influentially elaborated by Holmes. See, for example, his comments in *International News Service v Associated Press*, 248 US 215 at 246 (1918).

[34] Hohfeld also underlined that many property rights are only limited and conditional. Thus, even a fee simple in land bestows rights which offer less than complete exclusivity, the owner's claims being potentially subordinate to such things as eminent domain, taxation, public easements, the law of nuisance and so on.

developed by Hale in his work on rate regulation and the valuation of shares. Indeed, nowhere was the legally constructed nature of property and property rights clearer than in relation to the new intangible forms of property—especially property in the form of income rights—which began to proliferate in the nineteenth century and which have since grown so enormously in importance. These forms of property do not generally concern direct rights over concrete things whose existence as objects is independent of law. Their underlying nature as bundles of rights is, therefore, always closer to the surface, as reflected in their original *in personam* character and the tendency to classify them as *choses in action*. In these cases the very objects of property themselves—'things'—have to be legally constituted, even if once constituted they tend to take on a reified life of their own. As James Penner says, 'there is n[o] world of 'things' out there all ready to be appropriated as property …., 'thing' in the context of property is …. a term of art'.[35]

III. CONSTITUTING AND SECURING FINANCIAL PROPERTY

Joint stock company shares, for example, were originally conceptualised not as autonomous property forms—as 'things' in their own right—but as (literally) shares in both the assets and contracts of companies, as a result of which they could not freely be transferred.[36] The processes whereby they were transformed into no-obligation, no-responsibility, no-liability, freely transferable, *rentier* income rights—shares not in the corporate assets but in the corporate surplus, receiving revenue in the form (if not at the level) of interest[37]—were highly complex, entailing more than just permissive legal endorsement of commercial practice and the 'private', contractual arrangements that it spawned. The rights comprising shares had to be reconfigured in very particular ways, something which involved radical legal intervention.

[35] James Penner, 'The Bundles of Rights' Picture of Property' (1996) 43 *UCLA Law Review* 711 at 807.

[36] As in part 'shares' in the contracts of companies, for example, shares were conceptualised as *choses in action* which were not, in principle, assignable at law. When it was asserted that joint stock company shares were to be transferable, Chief Justice Best explained (in the days before they were almost invariably fully paid-up), it usually purported to mean 'that the assignee was to have all the rights of the assignor, and to take upon him all his liability'. In fact, Best argued, 'the assignee [could] join in no cause of action that accrued before the assignment', for such rights of action continued to lie with the assignor, who notwithstanding his withdrawal from the company 'remain[ed] liable for every debt contracted by the company before he ceased to be a member', *Duvergier v Fellows* (1828) 5 Bing 248. For similar observations, see John George, *A View of the Existing Law Affecting Unincorporated Joint Stock Companies* (London, McDowall, 1825). More generally, see Paddy Ireland, 'Capitalism Without the Capitalist: the Joint Stock Company Share and the Emergence of the Modern Doctrine of Separate Corporate Personality' (1996) 17 *Journal of Legal History* 41.

[37] As Marx explained, the return on shares is received in the form if not at the level of interest: as a reward for simply owning money.

Their establishment as autonomous, freely transferable property forms was a complicated and problematic judicial and legislative process, from which they emerged as decidedly circumscribed but nevertheless very valuable bundles of rights. It involved their gradual severance from obligation and responsibility (asset partitioning, separate corporate personality, limited liability, recognition of ever more ways in which they could become fully paid-up), the erosion of their original contractual, *personam* nature (free transferability) and a range of legal and other forms of protection for their passive holders. It was also, inevitably, a highly controversial and fundamentally *political* process, involving the promotion of very particular interests and significant derogations by both legislature and courts from traditional notions of contractual obligation and individual responsibility.[38] In the words of the *Edinburgh Review*, the acts of parliament which offered corporate (and other) privileges to companies 'qualif[ied] and dispense[d] with the general law'.[39]

Moreover, because of their essential nature as intangible income rights and the *rentier* nature of their owners, corporate shares were, as they still are, unusually vulnerable forms of property. This vulnerability stems in part from the so-called 'agency problems' associated with absentee ownership but more particularly from the source and nature of their value. The value of corporate shares is derived not from their concrete properties as physical objects or from the value of the tangible assets owned by the companies concerned but from their anticipated future earning power, from a capitalization of the dividends which are expected to accrue to them in the future.[40] Thus while the value of the tangible assets of companies tends to remain relatively stable, at least in the short term, share values often fluctuate alarmingly.[41] Indeed, there are frequently huge gulfs between the market and book values of companies, between the value of their tangible assets and their value in the stock market. In 1996, for example, at a time when

[38] The judicial and legislative opposition to joint stock companies and to the idea of freely transferable shares was deep-rooted and passionate, see Ron Harris, *Industrialising English Law: Entrepreneurship and Business Organization, 1720–1844* (Cambridge, Cambridge University Press, 2000). The same was true of limited liability. The basis of the law of partnership, explained the barrister Edward Cox in 1856, was 'that there is a moral obligation, which it is the duty of the laws of a civilised nation to enforce, to pay debts, perform contracts, and make reparation for wrongs'. Limited liability was 'founded on the opposite principle … permit[ting] a man to avail himself of his agent's acts if advantageous to him, and not to be responsible for them if they should be disadvantageous; to speculate for profits without being liable for losses; to make contracts, incur debts, and commit wrongs, the law depriving the creditor, the contractor, and the injured, of remedy against the property or the person of the wrongdoer, beyond the limit, however small, which it may please him to determine his own liability', Edward Cox, *The Law and Practice of Joint Stock Companies* 1st edn, (London, Law Times Office, 1856).

[39] 'Private Bill Legislation', *Edinburgh Review* (January, 1855) 152.

[40] *Short Brothers v Treasury Commissioners* [1948] AC 534; See Ireland *et al*, above n 26.

[41] For example, as I write the cable company Telewest is in the news. Its market value at the close of 2002 was about £44 million, having declined from a high a few years ago of about £7 billion, see *Guardian* 19 December 2002 26.

IBM's total market capitalisation was $70.7 billion, Microsoft's was $85.5 billion even though the company owned a mere $930 million in plant, equipment and property compared to IBM's $16.6 billion.[42] One consequence of this, as history has repeatedly shown and as we have seen only too clearly recently, is that shares are a form of property unusually vulnerable to manipulation, fraud and speculation.[43] As a result their security and integrity requires a wide range of further, costly, legal and other interventions aimed at eliminating deceit and swindling, and at trying to ensure that their market values reflect with reasonable accuracy their income-generating potential. These have been forthcoming in large part because of the economic and political power of their owners. Some have occurred within company law: as recently as 1980 a whole new tranche of provisions aimed at securing the integrity of corporate shares were enacted in what is now Part X of the Companies Act 1985. But the most significant interventions have occurred outside company law as traditionally drawn, with the development from the nineteenth century of modern accounting practices and standards and, more recently, of what has come to be called securities regulation or capital markets law, a legal realm in which the public presence of the state is only too apparent.[44] With its detailed stipulations and powerful

[42] See Jeremy Rifkin, *The Age of Access* (London, Penguin, 2000) 50–51. As Rifkin points out, this gulf is widening still further as a result of the growing importance of intangible, intellectual property forms. On this development, see text below.

[43] 2002 was a bumper year for corporate disasters and scandals, seeing major companies collapse at an unprecedented rate. Five of the biggest ten bankruptcies in American corporate history occurred in 2002 (Enron, WorldCom, Global Crossing, United Airlines and Conseco). The continuing scandal surrounding the Harvard Institute of International Development (HIID) and the economist Andrei Shleifer is also worthy of note, not least for the light that it casts on the processes of privatisation and on the mind-set of the privatisers and certain corporate governance 'specialists'. The HIID was employed by the US Agency for International Development to advise the Russian government on economic 'restructuring', receiving over $40 million for doing so. While directing the project and dispensing advice, however, Shleifer invested hundreds of thousands of dollars (directly and indirectly through family members) in Russian companies. Objecting to these 'activities for personal gain by personnel placed in a position of trust', USAID terminated its contract with HIID, and actions for damages were launched by both the US government and by an American mutual fund company against Harvard, Shleifer, and another HIID employee, Jonathan Hay, alleging, *inter alia*, fraud and breaches of contract, including violations of conflict-of-interest clauses. The investments, the government argued, undercut the project's mission of providing fair and unbiased advice on the privatisation of the Russian economy. Harvard removed Shliefer from the project, but retained him as an economics professor. It also disbanded the HIID, though it rapidly reappeared, along with most of its staff, under a new name. The suit with the company was settled quietly by the University (which had originally described it as 'baseless') in November 2002. The suit with the government, who are claiming triple damages of $102 million under the False Claims Act, is proceeding towards trial. See Janine Wedel, *Collision and Collusion: The Strange Case of Western Aid to Eastern Europe* (New York, Palgrave, 2001) ch 4. The progress of the lawsuits can be traced in the pages of *The Harvard Crimson*. One would like to think that the title of Shleifer's co-authored book, *The Grabbing Hand: Government Pathologies and Their Cures* (Cambridge Mass, Harvard UP, 1998) was intended to be deliberately ironic. He has recently been appointed an inaugural fellow of the European Corporate Governance Institute.

[44] The origins of securities regulation can be traced much further back into history, see Stuart Banner, *Anglo-American Securities Regulation: Cultural and Political Roots, 1690–1860*

administrators, this is a body of law which is ineradicably public and regulatory in nature.[45]

The vulnerability of intangible property in the form of income rights such as shares and bonds, however, extends well beyond the risks associated with fraud and speculation. As their value is derived from anticipated future revenues, they are potentially affected, individually and collectively, by any changes in the productive (and other) processes responsible for generating the income streams concerned. They are unusually vulnerable, in other words, to everything from changes in trade and labour regimes, to changes in financial and taxation policies, to changes in the balance of political and class power, to changes in particular markets as well as in general economic conditions—shares especially so, in that they provide no 'right' to an income, merely a 'privilege' to receive one in certain circumstances. As Robert Hale noted, the owners of these property forms are reliant upon large numbers of people 'conduct[ing] themselves in a manner advantageous to [them]' and continuing to do indefinitely into the future.[46] Their financial integrity therefore requires an extraordinarily wide range of legal and other regulatory activity aimed at preserving the general social and productive arrangements and conditions responsible for ensuring a continuing flow of money (interest).[47] In short, they are high-maintenance property forms which demand remarkable levels of general economic and social 'regulation'. Indeed, as they have acquired greater social significance—not least in relation to the provision of pensions for the middle classes in wealthier countries[48]—and as they have become increasingly international and mobile in character as investors scour the earth in search of better returns, the political pressure and demand for extensive support from a wide range of agencies at both the national and international levels has grown. These are property forms in increasing need of a very particular (new) world order.[49]

(Cambridge, Cambridge University Press, 1998), though it was not until the passing in 1933 of the Securities Act in the US that it began to grow in prominence and scope. Not until 1986 was legal provision finally made for a comprehensive system of securities regulation in the UK, with the enactment of the Financial Services Act and the establishment of the Securities and Investment Board.

[45] See Ben Pettet, *Company Law* (Harlow, Pearson Education, 2001) 6–9, 331 *et seq*.
[46] See Robert Hale, 'Value and Vested Rights' (1927) 27 *Columbia Law Review* 523. In Hale's terms, the value of these property forms is dependent upon a whole network of 'vested rights'.
[47] In this context, it is perhaps worth noting that the word 'credit' has its origins in the Latin *credere*, meaning to believe.
[48] Though it should not be forgotten that even within these countries ownership of shares and other financial property forms remains highly concentrated. By 1995, for example, more American adults (51.3 million) owned financial assets than ever before, but 71 per cent of households still owned no shares at all or held less than $2000 in any form, including mutual funds, popular employer-sponsored retirement '401k' plans and traditional pensions, see Jeff Gates, *The Ownership Solution* (Reading, Mass., Addison-Wesley, 1998) 4.
[49] It is also worth noting that the rise of the private pension has created growing pressure for secure and profitable outlets for money capital investment, something which has in turn increased the pressures for the privatisation, commodification and marketisation of anything and everything.

For confirmation, one only has to look at what has been entailed in the privatisation and marketisation processes taking place in the so-called 'transition' economies of Eastern Europe and elsewhere,[50] and at the nature and extent of the policy demands made by institutions such as the IMF and World Bank as pre-conditions of providing governments with the financial assistance they need in order to continue to service their debts. Indeed, in these contexts, the social relations which underlie these reified and fetishized property forms—and the fact that they represent private claims on what are fundamentally *social*, productive activities—are apt to come to the surface. It is only too clear, for example, who suffers when the IMF and World Bank, in the interests of interest, demand 'adjustments' and 'austerity programmes' involving reductions in social spending, privatisation of resources, export-led growth, wage freezes, the removal of subsidies to domestic industry, and so on. In Argentina, for example, one of the latest countries to hit the headlines, malnutrition amongst children has risen sharply since the government defaulted on its $141 billion debt and devalued the peso, notwithstanding the fact that Argentina is one of the world's biggest producers (and fourth largest exporter) of food. The country has just defaulted on a further $805 million debt with the World Bank and it is almost certain that any new deal with the IMF on debt repayments will be conditional on further 'austerity' measures ensuring that it is once again the poor who are hit hardest.[51]

[50] For an examination of the many social, political and legal pre-conditions of secure *rentier* investment, see Rafael La Porta, Florencio Lopez-de-Silanes, Andrei Shleifer & Robert Vishny, 'Investor Protection and Corporate Governance', 2000 *Journal of Financial Economics*. For critical accounts of what 'privatisation' and 'transition' have actually entailed in Russia, see Stephen F Cohen, *Failed Crusade: America and the Tragedy of Post-Communist Russia* (New York, Norton, 2000); and Peter Reddaway & Dmitri Glinski, *The Tragedy of Russia's Market Reforms* (Washington DC, US Institute of Peace, 2001). As Cohen makes clear, most academic (and journalistic) accounts of 'transition' in Russia in the 1990's represented a triumph of ideology over scholarship. 'In my own experience', he writes, 'the words *collapse*, *disintegration*, *tragedy* come more readily to the minds of most Russians than does *transition* ....', 44. 'Looting', he suggests, would be a more apposite description than 'privatization'. For an assessment of the gender impact of market-centred economic reform, see Rittich, above n 3.

[51] See *Guardian*, 11, 24, 25 November 2002. The latest default followed tough negotiations by the government with the IMF on a 'rescue package'. Among the thorniest issues for resolution were laws to protect foreign investments and businesses, and cuts in public spending. The link between (the payment of interest on) financial property and the exploitation and suffering of the poor, so evident in the financial crises which have afflicted countries such as Argentina, Brazil, Mexico, Russia, etc, is hardly new. Thomas Pakenham recounts how, when Ismael Pasha, Khedive of Egypt, defaulted on the interest due on his foreign loans in 1880, the European powers established an international commission to protect the bond-holders who had made a fortune lending money to him at extortionate rates of interest. Led by Britain and France, they moved in to depose Ismael and 'to rescu[e] Egypt—and the bond-holders'. The interest remained extortionate, but 'the real sufferers were the fellahin, the Egyptian peasants, who paid the interest with or without the encouragement of the courbash.' The courbash is a whip made of buffalo-hide. See Thomas Pakenham, *The Scramble for Africa* (London, Abacus, 1992) 78, 124, 128.

Put simply, then, financial property has become the principal mechanism whereby the rich extract wealth from the poor, prompting Wade Mansell to argue that Western law has performed the miracle of enabling the 'rich, developed, and often ex-colonial states … to … continue extracting wealth from the poorest countries' in an era of apparent decolonialisation. Gun-boat diplomacy and costly imperial administrations have largely (though not entirely) been replaced by other, less militaristic, less politically-overt, *legal* modes of 'regulation'.[52] In this context, it is deeply paradoxical that the recent rise of neoliberalism and of allegedly 'deregulated', 'free' markets and 'free' trade should be so closely associated with the growth and spread of intangible, financial property forms whose very constitution requires radical legal intervention and the maintenance of whose integrity and value demands the continuing management of so many spheres and aspects of social life. The reality is that preservation of the social processes which enable all money to act as capital and to command interest—and, through that, to lay claim to part of the social product—entails extraordinarily far-reaching legal and other 'regulation', not least because the processes often require not freedom, democracy and a shrinking of the state, but extended regulation, overt coercion and repression, commonly threatening not only the dignity but the very survival of many the world's poorest.[53] All in defence of what Hale referred to as the 'right to squeeze an income out of the community'.[54]

IV. INTELLECTUAL PROPERTY: THE NEW ENCLOSURES?

In the years since Hale wrote, intangible property forms have, of course, continued to proliferate, most notably with the rise of new types of intellectual property. As Jeremy Rifkin says, 'if the industrial era was characterized by the amassing of physical capital and property, the new era prizes intangible forms of power bound up in bundles of information and intellectual assets'.[55] With the emergence of the so-called 'knowledge' economy',[56] corporations—striving to increase profits in order to serve financial property—are seeking to construct all kinds of knowledge as patentable

[52] See Wade Mansell, 'Legal Aspects of International Debt' (1991) 18 *Journal of Law & Society*, 381. It goes without saying that these processes often pose a serious threat to meaningful democracy. As long ago as the 1820s Eduardo Galeano, commenting on the terms of a British loan to Argentina, wrote of the 'bars [put] around free nations' by 'such usurious operations', cited in Naomi Klein, *Fences and Windows* (London, Flamingo, 2002) xviii–xix.

[53] See Wade Mansell, Belinda Meteyard & Alan Thomson, *A Critical Introduction to Law* 2nd ed, (London, Cavendish, 1999) especially chs 9 & 10.

[54] Hale, above n 10 489.

[55] Rifkin, above n 42 30.

[56] A term which, like the term 'information society', is often misleadingly used to suggest that we are witnessing a shift to a form of capitalism which is radically different from its predecessor—indeed, in some versions to a form of society which is somehow beyond capitalism—rather than a radical extension of capitalist social relations.

and, therefore, as potential 'property' capable of enclosure. Patents are now granted not only for machines but for life forms, as, at the behest of corporate interests, everything from microorganisms to genes and molecules, to indigenous knowledges and medicines, to human cells and proteins has been deemed patentable.[57] The recent development of intellectual property law has thus provided numerous examples of the social and political processes whereby objects of property and property rights are legally constituted. Indeed, recent years have seen not only major extensions of the 'things' deemed patentable and capable of becoming 'property', but concerted corporate efforts to strengthen and extend the international protection of the intellectual property rights which have resulted, most notably in the form of the Agreement on TRIPs (Trade-Related Aspects of Intellectual Property Rights) forged under the auspices of the WTO during the Uruguay round of GATT.[58]

The recent expansion of intellectual property thus further underlines the socially constructed and contentious nature of property and property rights. As Christopher May says, intellectual property is legally constituted 'not to protect something previously existing, already recognised as property, but rather to protect certain current interests', interests whose position is, of course, immeasurably strengthened by being deemed 'property'.[59] Moreover, like Hale's work on rate regulation,[60] analysis of these property forms reveals the importance of their precise legal constitution—their exact rights-content—to their value and distributional effects: the value of intellectual property rights varies greatly according to the exact nature of the

[57] The development of IPRs in gene technology illustrates how the products of nature can be privately appropriated. Holders of intellectual property rights, David Campbell & Sol Picciotto explain, 'not infrequently consider themselves to be owners not merely of bundles of rights but of an asset, often conceived as a functional physical object ...'. They report how an IP lawyer who had worked for a corporation engaged in genetic patenting informed them that: 'Pharmaceutical companies do indeed wish to retain patent protection because of the control that it gives them. *It is their molecule* and they have the rights to that control' (their emphasis), see 'The Antinomies of Intellectual Property: Rationing, Regulation, Remedies and Revenues', paper delivered to the W G Hart Workshop, IALS, London, July 2002.

[58] Under the TRIPS agreement, the 144 members of the WTO are obliged to give effect to a set of minimum rules and principles concerning a wide range of IPRs. For a summary and assessment, Peter Drahos & Ruth Mayne (eds), *Global Intellectual Property Rights* (Basingstoke, Palgrave Macmillan, 2002). The agreement has proved highly controversial, particularly among developing nations. According to Vandana Shiva, 'TRIPs was not negotiated by GATT members. It was imposed by MNCs who used the US government to force it on other members ... It is the imposition of values and interests of northern MNCs on the diverse societies and cultures of the world', see *Protect or Plunder?: Understanding Intellectual Property Rights* (London, Zed Books, 2001) 95–96.

[59] Christopher May, *A Global Political Economy of Intellectual Property Rights: The New Enclosures?* (London, Routledge, 2001) 16.

[60] See Robert Hale, *Valuation and Rate Making* (New York, Columbia University Press, 1918). Hale pointed out that in these cases the value of the property concerned was a direct reflection of the rates set by the commissions, see Hale, 'Economic Theory ....', above n 7 194–95; see also Duxbury, above n 6 427.

monopoly rights (the amount of coercive power to withhold) legally attached to them. Take, for example, patents. Until relatively recently, the patent laws of many countries, Canada and Mexico for example, gave manufacturers shorter monopolies on newly patented drugs than did the equivalent laws in the US. The result was that drugs in Canada and Mexico were cheaper. However, the extension of higher American standards in the name of free trade has meant that instead of being able to compete on favourable terms with American companies, Canadian and Mexican companies have to move to the longer US patents. This was celebrated by the Vice President of one company. 'One of the preconditions of research-based pharmaceutical companies', he argued, was the development of 'first-rate patent laws as reflected in the standards of NAFTA'. But the distributional effect has been increased profits for drug companies and significantly higher drug prices and cutbacks for the Canadian NHS.[61] Here, health as well as wealth distribution is at issue.[62] In short, the TRIPs agreement represents the globalisation of US-style patent laws to the benefit of a small number of corporations (and their shareholders) but to the detriment of nearly everyone else. One of its 'clear effects', writes Peter Drahos, is to 'set up a flow of revenue from the less developed to the more developed, thereby contributing to structural inequality around the world'. It is hard not to conclude, he argues, that through the rules of intellectual property 'the rich have found new ways to rob the poor'.[63]

The example of patents also shows, as Sir Arnold Plant pointed out many years ago, that one of the peculiarities of property rights to 'things' like patents and copyrights is that 'they do not arise out of the scarcity of the objects which become appropriated': knowledge is not depleted by shared use. On the contrary, their scarcity—and the markets they create—are 'deliberate creation[s] of statute law': property rights in patents and copyrights generate a scarcity 'which could not otherwise be maintained'. As a result, whereas one might normally expect state action to be directed at preventing the raising of prices, 'in these cases the object of the legislation is to confer the power [to] rais[e] prices'. Legislators place owners 'in a position to secure an income from the monopoly conferred … by restricting the supply in order to raise the price'.[64] Scarcity in these contexts is a legal

[61] See Barbara Garson, *Money Makes the World Go Around* (London, Penguin, 2001) 51. The prices rose by about a third.
[62] See Krugman, above n 1; Drahos & Mayne, above n 58.
[63] Peter Drahos, 'Introduction' in Drahos & Mayne, above n 58 1. For a similar view, arguing that rather than being 'a technical issue of legal refinement, the current global regime for intellectual property serves quite specific interests', see May, above n 59 12.
[64] Sir Arnold Plant, 'The Economic Theory Concerning Patents for Inventions', in Arnold Plant, *Selected Economic Essays and Addresses* (London, Routledge Kegan Paul, 1974) 36. None of this is to say that the traditional justifications of rights such as patents—that scarcity needs to be artificially created in order to create much needed incentives—is not without some validity. See Sol Picciotto, 'Defending the Public Interest in TRIPS and the WTO', in Drahos & Mayne (eds), above n 58 224.

construct, the fences erected being, to use Naomi Klein's imagery, virtual (and legal) rather than real.[65] The result is that the income collected by the property owner in these situations is 'not something created by that owner, but something squeezed out of others by a law-made pressure'.[66] These too are property forms which are not only legally constituted—'intellectual property rights are a form of government regulation of the market'—but property forms in continuing need of far-reaching regulatory support. The TRIPS agreement, Drahos suggests, is the outcome of a process of 'international regulatory capture' by corporations within the WTO.[67]

V. PROPERTY, PRIVATE GOVERNMENT, AND THE MYTH OF DEREGULATION

For Robert Hale the socially and legally constructed nature of property and property rights underlined the extent to which in modern capitalism markets are themselves in crucial respects legal constructs, the products of legal 'regulation', of decisions by the state to create rights to invoke the power of law to exclude. They require, in other words, the legal constitution of property rights and in many cases the legal constitution of the very 'things' which are to be the objects of property and exchange relations. Many capitalist markets are not and can never be, as many contemporary supporters of neoliberalism claim or imply, purely 'private' and, by implication, naturally arising, unregulated or potentially unregulated, spheres of voluntary exchange.[68] As Hale pointed out, even when the courts 'accord to private parties the initiative in the creation of some of the rights and duties which they, the courts, enforce', they are 'not ousted from jurisdiction'.[69] There are,

[65] Klein, above n 52, xxiii.

[66] Hale, 'Economic Theory ….', above n 7 215. Hale considered all incomes to be the result of coercion. 'All incomes', he argued 'in the last analysis, whether derived from ownership of property or from personal services, are not 'products' created by the recipients; they are payments derived from the rest of the community by exertion of some sort of pressure. To say this is not to condemn the exertion of such pressure; it is the only means a man has, under present arrangements, and perhaps under any workable scheme of things, for keeping alive. But to appreciate that each income is the fruit of coercion exercised either with the aid or without the hindrance of the government, is to deny that a coercive governmental alteration of the size and distribution of incomes is necessarily wrong. No particular income is sacred merely because it is the result of 'free bargaining'. The coercion in which it originated is not necessarily less 'artificial' or 'arbitrary' or even 'confiscatory' than that which would take it away in whole or in part. The justification of each income must rest on some other ground than that the recipient has 'produced' it'. Hale, 'Economic Theory ….', above n 7 216–217.

[67] Drahos, above n 58 1–4.

[68] For a discussion of how this mode of analysis has been used by neoliberal 'nexus-of-contracts' theorists to try to assert the private, contractual nature of the public corporation, see Paddy Ireland, 'Recontractualising the Corporation: Implicit Contract as Ideology', in David Campbell, Hugh Collins & John Wightman (eds), *Implicit Dimensions of Contract* (Oxford, Hart, 2003), 255.

[69] Hale, 'Law Making ….', above n 14 452.

therefore, crucial senses in which law is inherently 'regulative' even when it is portrayed as, and claims to be, not interfering or regulating. Only by naturalising property rights and bargaining rules can it be thought otherwise. As a result, Hale and the realists argued, the public-private, state-market dichotomies which lay at the heart of early twentieth century laissez-faire ideology and which continue to lie at the heart of early twenty-first century neoliberalism do not stand up to critical scrutiny.

For Hale the relational and exclusionary nature of property rights also underlined that markets are not the spheres of freedom claimed by many of their advocates but are, rather, intricate, legally constituted networks of mutual coercion, 'complicated network[s] of restraints, imposed in part by individuals, but very largely by the government itself at the behest of some individuals on the freedom of others …'.[70] In 'protecting property', the law is 'intervening to restrict what would otherwise be the liberty of the non-owner'.[71] Even in allegedly non-coercive, 'free' market economies, therefore, the laws of property and contract intervene to impel some to fulfil what others claim from them and to respect the claims of others to be free from intrusion. By entitling their possessors to dictate the terms and conditions with which others have to comply in order to have access those things designated property—or, as in the contemporary corporate context, with which they have to comply in order to prevent firms from moving their property and jobs elsewhere—private property rights give coercive 'governing power' to private individuals. Like the systems advocated by professed upholders of laissez-faire in the late nineteenth and early twentieth centuries, therefore, the systems advocated by contemporary neo-liberal supporters of privatisation and deregulation are 'in reality permeated with coercive restrictions of individual freedom'.[72] Markets can never be coercion-free zones, for the freedom of the market is the freedom of individuals and groups to coerce one another.[73]

Moreover, as Hale stressed, in markets the stronger parties do not usually rely on their own force but on the force applied by the courts and other state agencies, 'first to exert pressure on the weaker to submit 'voluntarily'

[70] Robert Hale, 'Economic Nationalism v Representative Government' (unpub ms.), cited Fried, above n 6 50.

[71] Hale, 'Economic Theory ….', above n 7 215.

[72] Hale, above n 10 470.

[73] Even in cases where law enforces particular rights and duties created by individuals, Hale argued, 'they are created (or modified or extinguished) by virtue of the power of mutual coercion (in the form of pre-existing rights) vested by the ordinary law in the two contracting parties. It will not do to say that the party to a contract is a voluntary agent merely. He makes the contract in order to acquire certain legal rights he does not now possess or to escape certain legal obligations with which he is now burdened. Were his liberty not restricted by these obligations imposed on him by the law and enforced in the ordinary courts, he might never submit himself to the new obligations of the contract. Thus in a sense each party to the contract, by the threat to call on the government to enforce his power over the liberty of the other, imposes the terms of the contract on the other', Hale, 'Law Making ….', above n 14 452.

to his terms, and then to enforce those terms after submission'.[74] In effect, officials 'carr[y] out the mandate of property owners', endowing the latter with the power to call on the governmental authorities.[75] All legal rights to the possession, use and exchange value of property, Hale therefore argued, are best and most accurately seen as delegations by the state to private individuals of a 'discretionary power over the rights and duties of others' which is not subject to direct democratic control.[76] These rights are a form of what he called 'private government', delegations of public authority to 'unofficial minorities',[77] something which prompted him to describe government as a private as well as a public phenomenon. 'In matters of everyday life', he wrote, 'our liberty is restricted by requirements laid down by those who have superior economic power. These stronger persons are not called rulers, or governors, nor are their dictates known as laws or ordinances, however great the pressure which forces obedience'. However, although the power of these economic superiors is 'not thought to be 'government' at all', and although 'liberty [is not] thought to be curtailed by it',[78] as a result of it individuals are in reality 'no less subject to state power in ostensibly non-regulatory, laissez-faire regimes than in openly regulatory ones'.[79] Indeed, in situations in which the structure of mutual coercion is heavily skewed, freedom of contract, as Friedrich Kessler observed, often enables people 'to legislate by contract and, what is more, to legislate in a substantially authoritarian manner without using the appearance of authoritarian forms'.[80] From this perspective, 'deregulation' emerges as an Orwellian euphemism for the extension of private governing power and its accompanying (financial and intangible) property forms, with all the extensive regulatory support that they increasingly demand. The other principal contemporary manifestations of neoliberalism, from privatisation to the introduction of market and quasi-market mechanisms to the GATT, also take on a rather different hue. As I write, Brazil has just elected a man of radical roots and views to the Presidency. The strictures imposed by the

[74] Hale, 'Law Making ….', above n 14 453.
[75] Robert Hale, 'Force and the State: A Comparison of 'Political' and 'Economic' Compulsion' (1935) 35 *Columbia Law Review* 149 at 198
[76] Hale, above n 75 149.
[77] Hale, 'Law Making ….', above n 14. He also emphasised that property was a governing power which was very unequally distributed, the existing constitution and allocation of property rights producing a distribution of wealth which was not only heavily skewed but extremely difficult to reconcile with any intuitive conception of justice.
[78] Robert Hale, cited Duxbury, above n 6 435.
[79] Mark Kelman, *A Guide to Critical Legal Studies* (Cambridge, Mass, Harvard University Press, 1990) 321–22. For confirmation one needs look no further than the enormous power wielded by 'investors', the owners of interest-bearing financial property forms, over the economic and social policies of the governments of many developing countries through the agency of organisations such as the IMF.
[80] Friedrich Kessler, 'Contracts of Adhesion—Some Thoughts about Freedom of Contract' (1943) 43 *Columbia Law Review* 629 at 640. Kessler's article appeared immediately after Hale's 'Bargaining Duress, and Economic Liberty', above n 14.

capital markets and the IMF (acting, of course, on behalf of the owners of intangible, financial property) will, however, make it impossible for Lula da Silva's government to implement the social and economic policies so badly needed by the poor of his country. He is not, of course, the first democratically-elected political leader to find himself in this position, nor will he be the last. The plight of the Brazilian poor is merely the latest example of what Hale referred to as the high price paid by the owners of little property for the maintenance of existing property values.[81] It is not difficult, then, to see why one writer describes the 'free' market and deregulation as 'something the rich world imposes on the poor'.[82] Or why another, writing of the United States, suggests that it has 'a socialist system—full of benefits—for the rich and free market capitalism for the working class'.[83]

VI. LAW AND THE DISTRIBUTION OF WEALTH

Robert Hale concluded that the choice faced by legal policy makers, legislative and judicial, is not between market-contractual freedom (private) and state coercion (public), or between regulation (public) and no-regulation/deregulation (private), for all societies are, inevitably, characterised by both volitional freedom and coercion. Every pattern of volitional freedom has a correlative, publicly constituted pattern of coercion; every lawful 'economic' power is, therefore, in certain crucial senses a political power, every economic inequality a political inequality. The invisible hand is invisible, it turns out, because it's not there. Policy decisions cannot, therefore, coherently be derived from the binary logic associated with the traditional freedom/coercion, public/private dichotomies of free market liberalism, or, indeed, from *a priori* notions of efficiency which presume (among many other things) that rights take a particular (alienable) form and have already been allocated.[84] As Hale and the early twentieth century critics of laissez-faire demonstrated, none of these ideas and distinctions are capable, in and of themselves, of shedding decisive light on the choices that face courts and legislatures trying to adjudicate over competing social claims. Nor can

[81] For an astute assessment of the problems faced by Lula, see Perry Anderson, 'The Cardoso Legacy', *London Review of Books* 12 December 2002 18. Hale himself observed in 1927 that the American Supreme Court, in its interpretation of the constitution, gave 'legislative bodies ... a far freer hand in making progress at the expense of vested rights' than [the] State Department [wa]s willing concede to weaker foreign governments'. Hale, above n 46 529. Little, it would seem, has changed.
[82] George Monbiot, *Guardian*, 29 October 2002.
[83] Derrick Bell, *Guardian* 3 December 2002.
[84] Among the influences on Hale was Richard T Ely who had earlier observed that there was a tendency among economists to treat private property as a 'natural' institution, see Ely, *Property and Contract in Their Relation to the Distribution of Wealth* (New York, Macmillan, 1914). Hale himself noted that the role of property law in wealth distribution was almost entirely ignored in standard economic texts, see Fried, above n 6 78.

policy decisions be coherently or justifiably derived from bald property claims based upon mistaken, naturalistic, reified conceptions of property—in which legal 'regulation' is presented as something externally-imposed by the state on the owners of pre-political property ('things'), rather than as something inherent in and constitutive of property and property rights themselves. Such assertions simply substitute for properly argued, wide-ranging, rational, policy debate. The critical questions for policy are not whether to back freedom or coercion, the public or the private, but rather which structures of rights, coercion and volitional freedom to adopt in particular contexts.[85]

None of this is to say that Hale was against private property *per se*. On the contrary, he saw himself as its defender. '[I]f property is not revised methodically by its friends', he argued, 'it is likely to be revised unmethodically by its enemies'.[86] He considered property rights to be *social* institutions, however, and, as such, to be mutable and contingent. They are, in his view, potentially valuable mechanisms for creating incentives and stimulating productive activity, and potentially valuable ways of regulating consumption.[87] The inequalities in wealth that tends to accompany them can often, if they are appropriately constituted, be justified in terms of the increased wealth that they generate and their contribution to individual autonomy and freedom. '[M]any of the inequalities of income [found in society]', he wrote, 'serve the useful function of stimulating production in such a way as to benefit others as well as the producers'.[88] But this is not always the case. There was, for example, in Hale's view, an important distinction to be drawn between property-for-use and property-for-power. 'When the owner of things uses those things himself', he argued, 'the legal arrangement which restricts the liberty of the non-owners promotes an interest of a very different sort from that promoted when the owner does not himself use the things owned'.[89] In the latter case, the function of property rights is not to promote the interests of owners in personal use of the things owned but to enhance their bargaining power and claim on an income from the community.[90]

[85] On this see Rittich, above n 3, 132–151, 160–61.
[86] Hale, 'Rate Making ….', above, n 7 216.
[87] See Hale, 'Economics and Law', above n 7 140 *et seq* for a discussion of taxes and incentives.
[88] Hale, above n 10 482.
[89] Hale, 'Economics and Law', above n 7 136–7, drawing on LT Hobhouse, 'The Historical Evolution of Property, in Fact and in Idea', in Bishop of Oxford (introduction), *Property: Its Duties and Rights* (London, Macmillan, 1915).
[90] Hale, 'Economics and Law', above n 7 137. Property rights, Hale reasoned, enabled owners to shape the terms upon which use of their property is rendered lawful. The terms imposed on workers, for example, 'include their surrender of all share of the title to the goods produced with the help of their labor in return for specified money wages. After production, the owner of the plant thus becomes the owner of the products and is then in a position to impose terms on those who would like to consume them'.

What Hale did object to, and objected to most strongly, was the circumvention of careful, context-specific, consideration of the difficult normative questions and policy choices surrounding the legal creation and disposition of state-backed private rights by the use of arguments based upon *abstract* notions of the 'natural' rights of property owners and freedom of contract. These he saw as attempts to avoid proper discussion of the policy issues involved in all their discomforting complexity; policy issues which need to recognise that the most we can ever achieve is a relative degree of freedom, and which involve questions not merely of productive efficiency but of justice and distribution of the social product. The courts, he suggested, 'usually assume that … economic consequences are justifiable simply because they are consequences of property rights which are supposed to be a fruit of legal equality; whereas in fact the property rights are part of a legal arrangement whereby the law curtails the liberty of different individuals in different degrees, and the justifiability of the particular arrangements depends on the justifiability of the economic results rather than the reverse'.[91] In formulating legal policy, then, distributive questions cannot be avoided because law 'endows some with rights that are more advantageous than those with which it endows others … [And] it is with these unequal rights that men bargain and exert pressure on one another…, giv[ing] birth to the unequal fruits of bargaining'. Different configurations and allocations of rights inevitably generate very different patterns of economic relationships.[92] The re-emergence of neo-liberalism has seen resource allocation increasingly presented as a function of 'the market', rendering it seemingly unnecessary for governments to have particularly clear or definite notions as to how incomes should be distributed. But the productive and distributional effects produced by markets are not the neutral and value-free outcomes of unregulated, private activity, the results of 'natural' laws of distribution. They are, rather, the products of the many contentious legal policy choices which create the network of pressure and counter-pressure that makes up the structure of mutual coercion. At present, 'some men have law-given rights over others which enable them to collect large incomes from the community, others have law-given rights which enable them to collect little or nothing at all'.[93] However things appear,[94]

[91] Hale, 'Economics and Law', above n 7 140.
[92] While potentially providing 'just a[s] strict protection of each person's property rights, and just as little governmental interference with freedom of contract', see Hale, 'Bargaining, Duress …', above n 14 628.
[93] Hale, 'Economic Theory …', above n 7 215. 'Most of our present distribution of wealth is the result of the relative power, latent or active, of various individuals and groups. The power itself is derived in part from law's more or less blind and haphazard distribution of favours and burdens, in the shape of powers over others and obligations to others', Hale, 'Law Making …', above n 14 455.
[94] Hale frequently emphasised that the distribution of wealth often appeared to be, and was commonly portrayed as, the product of voluntary agreements made by private individuals, see Kennedy, above n 6 86.

state and law, politicians and judges, are inevitably deeply implicated in resource allocation and the distribution of wealth in society.

This prompted Hale to call for a frank examination of the functions of private property in order to 'work out a body of law for the revision of property rights where they need revision, and for their preservation where they need preserving'. He conceded that the results might be radical but argued that if this were so 'it would be because on a piecemeal and candid review, many of the incidents of property would prove themselves without justification'.[95] He conceded also that such an examination would undoubtedly impact on vested rights, but observed that 'the more perfect the guaranty against any loss of property value, the greater the burden on the non-propertied classes'.[96] More than ever, we urgently need debate about how rights should be configured and allocated, with attention being given to questions of social welfare, income distribution, and the channels into which industry worldwide should flow, as well as to questions of productive efficiency. 'The owner of every dollar', Hale observed, 'has, by virtue of his law-created right of ownership, a certain amount of influence over the channels into which [productive activity is directed]'. Today, it is more obvious than ever that those with the most money exercise the most control over these channels and that the less pressing needs of the rich have far greater influence over them than the more pressing needs of the poor.[97]

Nor should we allow these debates to be curtailed by simple claims of property right and strident railings against 'regulation'. The rights of property, Hale argued, should 'prove themselves on examination in each particular case and submit to modification where a balance of gain for the public can be reasonably expected'.[98] Ultimately, economics alone can give no decisive answer to questions of wealth distribution and the proper channels of industry. It can trace the different economic consequences of different courses of action, 'but once it has made the issues clear, the final judgement [on them] is an ethical one'.[99] In an era of rampant privatisation in which, with the growth of the life-science industry and gene technology, attempts are being made to extend the coercive power of private property still further, Hale's insights are more relevant than ever, as is his call for a frank and candid examination of private property rights, markets and their functions.[100]

[95] Hale, 'Rate Making ...', above n 7 216.
[96] Hale, above n 46 528.
[97] Hale, above n 10 490–91. For similar arguments made in the context of the constitution of IPRs, see Drahos, above n 63; Picciotto, above n 64.
[98] Robert Hale, 'Political and Economic Review' (1924) 10 *American Bar Association Journal* 51.
[99] Hale, 'Economic Theory ...', above n 7 225.
[100] For Hale the kind of analysis that was required to determine appropriate rates in public utilities needed to be extended to all property forms. In the rate regulation cases, he argued, the courts were, in effect, working out standards for determining both the precise nature of the property rights of utility shareholders and what it was proper for them to receive (in income terms) from their property ownership. Potentially, Hale believed, it provided a model for the revision of property rights as a whole, see 'Rate Making ...', above n 7 213.

At present, in order to preserve certain property forms and their economic value (corporate securities, government bonds, IPRs and the like), law intervenes in all aspects of social life, and intervenes on a massive scale, in ways which are often ethically indefensible.

VIII. THE ROLE OF THE LEGAL ACADEMIC

'To the conventional eye, governing power is invisible save when exerted by public officials wearing the authentic trappings of the political state.'[101] One of the roles of legal academics is to make visible the other forms of legally constituted governing power, to expose the legal bases of the structure of *private* economic power, and to bring to the fore the particular interests which are elevated by the often hidden policy choices embodied in law. This entails, among other things, dispelling the widespread view of markets as naturally arising, potentially unregulated, private phenomena. Markets are not a product of nature. On the contrary, they are to a large extent 'legally constructed instruments, created by human beings hoping to produce a successful system of social ordering ... there is no opposition between 'markets' and 'government intervention''. Like all such instruments, they should be 'evaluated by asking whether they produce our social and economic goals'.[102] It also entails dispelling the common-sense, naturalistic, reified understanding of private property as an owner's pre-political, exclusive right to a 'thing', with its accompanying presumption against external state intrusion. Like markets, property rights are social institutions created and protected by the state and law, often at considerable cost. They too need to be justified in terms of their social use and value.[103] The key policy questions are not then simple matters of freedom versus coercion, of 'regulation' versus 'deregulation'. Both coercion and government regulation, particularly in the form of the constitution and enforcement of property rights, already restrict liberty and do so drastically and unequally. The key policy questions concern the nature and effects of government intervention. At present, markets are constituted by law in a manner which 'differentially and asymmetrically empowers groups over the fruits of cooperation in production'.[104] Sometimes one needs not merely to limit the power of

[101] Robert Hale, cited by Duxbury, above n 6 435.
[102] Cass Sunstein, *Free Markets and Social Justice* (Oxford, Oxford University Press, 1999) 384.
[103] It is worth noting that Hale thought this particularly true of income rights. 'Like all special privileges of a pecuniary nature', he wrote, 'we seek to allow income from ownership to the extent, and to the extent alone, that it conduces to some social purpose .... We aim to measure the privilege by the end to which it is supposed to conduce ... [and] to leave the owner so much income only as is thought socially desirable ...', Hale, cited Samuels, above n 6 269–70. Hale did not think that all income rights should be abolished but that their social value should be assessed in a context-specific manner, see Hale, 'Rate Making ...', above n 7 214–15.
[104] Kennedy, above n 6 83.

government to impose arbitrary restraints on individuals, but to invoke its power to restrain the more powerful from imposing arbitrary restraints on the less powerful—not just in the interests of greater equality but in order to safeguard liberty itself.[105]

Exposure of the invisible threads of power is not only a legitimate task for academics to undertake, but one which is growing in importance, for Hale's claim that we simply cannot escape distributional questions is, surely, more relevant than ever in the world in which we live, a world characterised not only by increasing social and productive interconnection (globalisation),[106] but also by a distribution of wealth, usually justified in terms of 'free' markets, which is becoming ever more skewed and ever more difficult to square with any intuitive sense of justice. It is also a task which demands a certain degree of critical separation from the world of commercial legal practice. For practising commercial lawyers are, inevitably and quite properly, partisan, representing clients with particular interests to protect, interests which often depend on the creation, extension and preservation of private property rights and the particular structures of mutual coercion that accompany them. These clients are very often best served by the concealment of the particularity of their interest, or by the representation of their interest as the interest of all. By contrast, one of the functions of legal academics is rigorously to analyse the world and to try to expose concealed or disguised interests. This is not, arguably, a task that they are currently performing with much distinction. Robert Hale called for 'the development of a true science of the economic aspects of the law', expressing the hope that the growth of such a science among lawyers 'might lessen the influence of crude economic ideas over the formation of other legal arrangements'[107] and encourage 'intelligent economic judgement[s]'.[108] He would, one suspects, be very disappointed with much of what presently passes for economic analysis of law.

[105] Robert Hale, 'Our Equivocal Constitutional Guarantees' (1939) 39 *Columbia Law Review* 563. For a discussion of the centrality to Hale's work of a positive (rather than a negative) conception of freedom, see Fried, above n 6; and Hale, *Freedom Through Law*, above n 6.
[106] For Hale's own recognition of the increasingly collective nature of productive activity in the modern world, see above n 10 603 *et seq*.
[107] Hale, 'Economics and Law', above n 7 142.
[108] Hale, above n 46 529.

Commentary on 'Property, Private Government and the Myth of Deregulation': Implicit Regulation and the Development of Regulatory Policy

ANDREW WHITTAKER

I. THE 'IMPLICIT REGULATION' THESIS

THIS COMMENTARY OFFERS some perspectives on one aspect of Paddy Ireland's chapter 'Property, Private Government and the Myth of Deregulation', focussed particularly on his argument that there is some regulation implicit in private law, particularly in the private law of property and the private law of contract. The commentary attempts to illustrate the various ways in which this implicit regulation presents itself, and argues that the existence of this implicit regulation means that, while the concept of 'free markets', and 'freedom of contract' convey important paradigms about different approaches to the development of markets, they should not be allowed to distort fundamental policy issues about the forms or degree of regulation which may be appropriate either to secure the better operation of markets as markets or for other public policy reasons, such as to inhibit the use of the financial system for money laundering. I offer these comments from the perspective of a lawyer involved in financial services regulation, and will also make some comments about what we can learn from financial services regulation about when it may be appropriate to replace the implicit regulation we inherit in our existing law with more explicit regulation designed specifically to achieve particular public policy objectives.

Paddy Ireland argues that the last two decades have seen the ambit and power of 'the market' considerably extended, and that the underlying supposition is that free markets are the best way to ensure the efficient allocation of resources and to maximise economic growth and wealth for the benefit of society as a whole. As a result, he suggests, much regulation is in practice designed not so much to replace market goals as to assist in their realisation, tackle market failures, or, in the public sector, substitute for their absence. He goes on to argue that the origin of current debates about the place and value of free trade, and the role of the state, is to be found in the law and economics movement in the nineteenth century, but that there was greater scepticism then about not only the capacity but also the notion of a free market. In his view, critically, since the private property rights and bargaining rules that constitute markets are legal constructs, both private property and markets are the products rather than the subjects of legal regulation. On this basis, he argues that legal regulation is both ubiquitous and inevitable. As a result, in his view, the creation and disposition of state-backed rights should be determined on the basis of the substantive issues and policy questions involved, rather than on abstract notions such as the 'natural' rights of property owners or 'freedom of contract'.

II. ASSESSING THE 'IMPLICIT REGULATION' THESIS

In my view, this analysis is right in a number of respects. It is right in recognising that property is often a legal construct—obvious examples from my own field are shares and unit trusts. The form of these property rights does indeed often involve policy judgements, for example, about the appropriate rights of shareholders, or unit holders, which have significant economic effects. Similarly, many sophisticated markets are indeed constituted by sets of bargaining rules—for example, the rules governing the respective roles of brokers and market makers on a stock market—which also have significant effects in regulating the operation of the market. 'Regulation' can be said therefore to be ubiquitous and inevitable, at any rate in the limited sense that the private law that enables people to give effect to their commercial intention itself involves a number of implicit public policy judgements, and that these may have grown up through custom and practice, through the cumulative effort of those developing a market, or through external public policy intervention.

It may be helpful to give some examples of different forms of implicit regulation inherent in our own commercial law.

There is, for example, some regulation of the *content* of bargains, through the mandatory mutual disclosure requirements in insurance contracts, the implication of terms into sale of goods contracts, the construction of exclusion clauses against the person seeking to rely on them, and the application of the law on force majeure, even when the parties have not expressly provided for it.

There are, similarly, a number of public policy limitations on the circumstances in which the courts will *enforce* bargains, including rules on capacity, for example of companies and minors, the treatment of illegal contracts such as gaming contracts, and the traditional constraints on usury and unfair credit bargains. In these cases the court declines to give effect to the agreement of the parties, because it is contrary to some overriding public policy requirement.

Further, competition law in its various forms, including the common law on restraint of trade, has applied implicit regulation to freedom of contract in the interests of the freedom and efficiency of the market as a whole, while copyright infringements are subject to a public interest defence similar to that in the law of confidentiality, in the interests of public policy.

These illustrations of implicit regulation forming part of our private law of contract support Paddy Ireland's thesis that freedom of contract, and freedom of the market, are of necessity hedged about by a series of legal rules designed both to give better effect to the intentions of the parties, and to limit the extent to which those intentions may be given effect. We accept these limitations as legitimate, because they are an established and accustomed part of our law. Indeed, we are often oblivious to them, compared with more overt forms of regulation. The implicit regulation thesis points out, however, that the 'regulation' implicit in the particular content at any given time of inherited property rights or bargaining rules may not necessarily be that which is best adapted for their current roles, and so where change is needed to such inherited rights and rules, we should not be deterred from considering change by an approach to what constitutes a free market which ignores implicit regulation.

III. ISSUES FOR CONSIDERATION

Instead, the real question in my view is always whether inherited regulation (which may be implicit, explicit, or a combination of the two) which is inadequate or out of date should be replaced or developed by a new regulatory initiative, designed either to achieve the original objectives better, or to achieve new objectives. This requires a complex judgement which brings into account the policy objective, how it is achieved at present, the options for achieving it in future, and the respective merits and demerits of the existing and possible alternative approaches.

The first issue for consideration seems to me to be how we identify *the scope of our authority*, as practitioners, courts or regulators, to develop regulatory policy through implicit or explicit regulation. Here, comparing and contrasting the methods of development of implicit and explicit regulation can be of assistance.

As *practitioners*, we influence the way in which the law develops by formulating standard terms of business, helping develop common understandings of the legal position in areas of uncertainty, and contributing our bit to the development of custom and practice. Our authority comes from our skill and influence, and the respect which others may accord it.

The *courts* develop implicit regulation on the basis of the implied will of the parties, the implied will of Parliament, and established and developing principles of public policy, taking inspiration on occasion from the principles of community law, the European Convention on Human Rights, or the decisions of overseas courts. Their authority comes from their jurisdiction to interpret and apply the law.

The basis for developing *legislative* regulation tends to be more overt. In the case of my own organisation, the Financial Services Authority, for example, its powers and objectives are set out by its enabling statute, the Financial Services and Markets Act 2000. This provides the FSA with rule-making powers to regulate the conduct of investment activities, so long as it does so for the purposes of one or more specified statutory objectives (taking into account a series of 'principles of good regulation'), consults publicly, and undertakes a cost benefit analysis.

The second issue in developing or replacing implicit regulation is how we ensure, in the context in which we are operating, that what we do *is no more restrictive of freedom than is needed to meet our objective*. Inherited implicit regulation manages this issue by adapting progressively and incrementally over a period. More wide-ranging or radical interventions need to be analysed from a range of different viewpoints, so that a range of options can be considered alongside their putative benefits and costs, including their potential impact on competition and innovation. They need to be capable of being reviewed if, in the event, they are more restrictive than needed, or they produce perverse results.

Finally, we need to *manage the transition* between one phase of regulation and another so that the development of the law does not unduly limit its predictability. Explicit regulation achieves this by explicit start dates and transitional provisions, against the background of the principle against retrospectivity. Court decisions find it more difficult to develop in this way, because of the fiction that they are merely declaratory of the existing law, though community law is more ready to accept that a court can determine the law for the future only. Practitioner-led developments flow naturally as new issues arise and business changes.

IV. CONCLUSION

In order to achieve public policy objectives, we need to identify the right combination of regulatory tools, whether relying on inherited implicit regulation, developing new implicit regulation, or devising some express external intervention. Workable regulation that achieves worthwhile goals depends on careful consideration of the objectives to be achieved, the available routes, and their potential effects. Paddy Ireland's analysis does a service to the development of regulatory policy by making it clear that the issues are much more complex than the headline conflict between 'free markets' and 'regulation'. A better understanding of their relationship will improve the prospects for better achievement of public policy objectives, including the better operation of 'free markets'.

Part 2

Contract Terms and their Interpretation

5

*The Intractable Problem of the Interpretation of Legal Texts***

JOHAN STEYN*

SOME MIGHT SAY that to speak of the intractable problem of interpretation of legal texts is an exaggeration. After all, unlike other professionals, a judge usually starts with the comfort that he has a 50 per cent chance of getting the answer to the question right. Moreover, he has the reassurance of Lord Reid's advice to judges that if your average drops significantly below 50 per cent you have a moral duty to spin a coin.

The centrality of the interpretation of legal texts is not always fully appreciated. Day by day, in Britain up and down the country, tribunals, lower courts, the High Court, the Court of Appeal, and the House of Lords, are concerned with the interpretation of a variety of legal texts ranging from wills, contracts, statutes, regulations, bye-laws, various types of 'soft laws' and so forth. It amounts to the preponderant part of the legal work of English judges, perhaps as high as 90 per cent. I would be surprised to hear that the position is significantly different in Australia. But the academic profession and universities have not entirely caught up with the reality that statute law is the dominant source of law of our time. The interpretation of legal texts is of supreme importance for a modern lawyer.

There are, of course, rules applicable to interpretation, some which are known by Latin expressions, such as *ejusdem generis, expressio unius est exclusio alterius,* and *noscitur a sociis.* These underlying rules of interpretation have a role to play. But, subject to what I will say about constitutional adjudication, ultimately interpretation does not generally depend on the application of rules. It is an art. And I would therefore like to start by taking a look at the subject in a more general way. But I am not putting forward a theory of interpretation. The subject is too elusive to be encapsulated in a theory. But as a

*A Lord of Appeal in Ordinary.

** This essay was first delivered as the John Lehane Memorial Lecture 2002 at the University of Sydney, 25 September 2002, and then published in *Sydney Law Review* [vol 25:5 2003]. It was reproduced for discussion at the Commercial Law & Commercial Practice Seminar, LSE, and again here with the kind permission of The Editors of the *Sydney Law Review.*

result of the work of legal philosophers, academic and practising lawyers, and judges it is possible to take stock of some modest insights.

I propose to discuss the interpretation of contracts, statutes and constitutional measures. But before I turn to these particular legal texts, I would put forward four general propositions which (if correct) go to the heart of interpretation.

First, it is a universal truth that words can only be understood in relation to the circumstances in which they are used. Adapting one of Wittgenstein's memorable examples, one can imagine parents telling a baby-sitter, who agreed to look after their five year old twins for some hours, that if the children become troublesome 'teach them a game'.[1] The parents return to find the baby-sitter playing poker with the children. Poker is a game. Did the context give a more restrictive colour to the word 'game'? Wittgenstein thought the answer was Yes. Judging by my own grandchildren I am not so sure.

The purpose of interpretation is sometimes mistakenly thought to be a search for the meaning of words. This in turn leads to the assumption that one must identify an ambiguity as a pre-condition to taking into account evidence of the setting of a legal text. Enormous energy and ingenuity is expended in finding ambiguities. This is the wrong starting point. Language can never be understood divorced from its context. In the words of Oliver Wendell Holmes a word is not a transparent crystal. The true purpose is to find the contextual meaning of the language of the text, ie what the words would convey to the reasonable person circumstanced as the parties were. In *Codelfa* Brennan J succinctly stated that 'the symbols of language convey meaning according to the circumstances in which they were used'.[2] Earlier this year in the *Royal Botanic Garden*s case your High Court reaffirmed this observation.[3] In his classic judgment in *Reardon Smith*[4] Lord Wilberforce illuminated this point. Speaking of contracts, Lord Wilberforce said that there is *invariably* a setting in which the language has to be placed. He made clear that the court is *always* entitled to be informed of the contextual scene of a contract. The same must apply to the interpretation of all legal texts. The failure to understand this fundamental principle of linguistic jurisprudence and legal logic has caused great injustices. An example in the field of wills is instructive. Consider the decision in *Re Fish: Ingram v Rayner*.[5] The testator gave his estate to 'his niece Eliza Waterhouse' during her life. The testator had no niece named Eliza Waterhouse. But his wife had a legitimate grand niece named Eliza Waterhouse and also an illegitimate grand niece of the same name. The illegitimate grand niece was living

[1] *Philosophical Investigations*, 1958, note to para 70.
[2] *Codelfa Construction Pty Ltd v State Rail Authority (NSW)* (1982) 149 CLR 337 at 401.
[3] *Royal Botanic Gardens and Domain Trust v Sydney City Council* (2002) 186 ALR 290, para 10.
[4] [1976] 2 Lloyd's Rep 621.
[5] [1894] 2 Ch 83.

with the testator and he was in the habit of calling her his niece. This was powerful objective evidence that the words in the will referred to her. With wringing protestations about the painful nature of their task, the Court of Appeal refused to admit the evidence. They held that there was no ambiguity. The illegitimate grand niece lost what had been left to her. What a grotesque result. The will could not be understood without knowing the context.

In the interpretation of legal texts the most frequent source of judicial error is the failure to understand the contextual scene of a legal text. Often judges are not provided with all the contextually relevant raw materials. The essential setting of a text may include in a contract case how a market works, in a breach of statutory duty case competing policy arguments, the structure of a complex statute, the historical development of legislation, and so forth. There is scope for the development of something like a *Brandeis* brief but carefully and concisely targeted to the relevant context.

The second proposition is that the aim of interpretation of a legal text, whether it be a private instrument or a public statute, must be to derive a meaning from its nature and contents. The mandated point of departure must be the text itself. The primacy of the text is the first rule of interpretation for the judge considering a point of interpretation. Extrinsic materials are therefore subordinate to the text itself. Often lawyers argue cases on the reverse hypothesis. Justice Frankfurter recalled the lawyer who said to the United States Supreme Court 'the legislative history is doubtful so I invite you to go to the statute'. Contextual materials must of course not be downgraded. On the contrary, the judge must consider all relevant contextual material in order to decide (a) what different meanings the text is capable of letting in and (b) what is the best interpretation among competing solutions. But the judge's task is interpretation not interpolation. What falls beyond that range of possible contextual meanings of the text will not be a result attainable by interpretation. There is a Rubicon which judges may not cross: principles of institutional integrity forbid it.

The third proposition relates to the generalisation that there has been a shift from literal interpretation to purposive interpretation. What is literalism? This is straightforward. The tyrant Temures promised the garrison of Sebastia that no blood would be shed if they surrendered to him. They surrendered to him. He shed no blood. He buried them all alive.[6] That is literalism. It has generally no place in modern law. On the other hand, it would be an over-simplification to say that there has been a homogenous shift towards a purposive interpretation of all legal texts. Much depends upon the particular text. A comparatively strict interpretation of a documentary credit issued in an international sale may be necessary because a third party (the bank) must be able to rely on a meaning gathered largely within the four corners of the text. In a network of contracts governing a construction project, parties ought

[6] This example is given in *The Works of William Paley* (1838 ed), v III, 60.

generally to be able to rely on the obvious meaning of the interlocking texts. Similarly, fiscal legislation may sometimes require a stricter approach than social welfare legislation. By contrast in a consumer transaction the purchaser of a fridge in a consumer sale may be entitled to a more generous interpretation of a right to reject a fridge which cannot make ice.

The fourth proposition I have already foreshadowed. Interpretation is not a science. It is an art. It is an exercise involving the making of choices between feasible interpretations. Structural arguments must be considered. Competing consequentialist arguments must be taken into account. Broader policy considerations may be relevant. Educated intuition may play a larger role than an examination of niceties of textual analysis. The judge's general philosophy may play a role. Ultimately, however, a judge must be guided by external standards in making his choice of the best contextual interpretation. He must put aside his subjective views and consider the matter from the point of view of the reasonable person.

I. COMMERCIAL CONTRACTS

Clarity is the aim in drafting commercial contracts but absolute clarity is unattainable. And it is impossible for contracting parties to foresee all the vicissitudes of commercial fortune to which their contract will be exposed. Moreover, and quite understandably, business bargains have to be struck under great pressure of time and events. Often the phenomenon of studied ambiguity obtrudes: the parties cannot resolve a particular difference but leave it to the court to settle the issue. It is therefore tiresome for judges to expatiate on the quality of draftsmanship of commercial contracts. Judges must simply do the best they can with the raw materials produced in the real world.

The common law does not in principle differentiate between the interpretation of a rudimentary cobbled-together contract and a sophisticated standard form contract; between the interpretation of a consumer contract and a commercial contract; or between the interpretation of a domestic and transnational contract. That is not, however, to say that in working out what is the best interpretation of a contract a court may not have to take into account, for example, a consumer as opposed to a commercial context, or the need for uniformity in international transactions.

In sharp contrast with civil law legal systems the common law adopts a largely objective theory to the interpretation of contracts. The purpose of the interpretation of a contract is not to discover how the parties understood the language of the text which they adopted. The aim is to determine the meaning of the contract against its objective contextual scene. By and large the objective approach to question of construction serves the needs of commerce. It is, however, less well suited to delivering practical justice in consumer transactions. There is much to be said for approaching commercial transactions and consumer agreements somewhat differently. This is already happening. One of the biggest modern developments in contract law has been the development of greater rights for consumers, notably in controls on exemption clauses and requirements of 'fairness' in consumer contracts. In England the principal impetus has been European directives for the protection of consumers. This has given rise to arguments that there should be two contract laws, one for consumer transactions, the other for commercial dealings.[7]

Two recent decisions in the House of Lords explored the extent to which the context may impress a meaning other than the obvious meaning on contractual language. In *Mannai Limited v Eagle Star Assurance Company Limited*[8] the issue was whether a contractual notice by a tenant to determine a lease was valid. The notice wrongly named the day upon which the tenant would do so as 12 January rather than 13 January. The majority held that the notice was valid. Essentially, they regarded it as wholly implausible that the tenant only wanted to terminate if he could do so on 12 rather than 13 January. Given this position the majority concluded that a reasonable recipient would have understood that the option was being exercised. In the context 12 meant 13. The minority held that the notice failed to conform to the requirements of the option reserved in the lease. As a member of the majority in *Mannai* I acknowledge that when judgments were delivered in the House of Lords Chancery practitioners hoisted a black flag over Lincoln's Inn.

The case of *Investors Compensation Scheme Limited v West Bromwich Building Society*[9] is important. The particular dispute can be put to one side. It is sufficient to say that by a majority of 4 to 1 the House of Lords upheld the conclusion of the judge that something had probably gone wrong in the drafting and reversed a ruling of the Court of Appeal. Lord Hoffmann, speaking for the majority, rejected the contention that judges cannot, short of rectification, decide on an issue of interpretation that parties had made mistakes of meaning or syntax. Lord Hoffmann observed that 'if one would nevertheless conclude from the background that something must have gone wrong with the language, the law does not require judges to attribute to the parties an intention which they plainly could not have had.' This was the ratio of the decision. However, in the course of his speech Lord Hoffmann also observed that the admissible background 'includes absolutely anything which would have affected the way in which the language would have been understood by a reasonable man'. This proposition upset the horses in the commercial paddock. Commercial judges vented their angst. Subsequently, in *BCCI v Ali* Lord Hoffmann explained that his observation only referred to anything which the reasonable man

[7] Appleby, *Contract Law,* (2001), chs 15 and 17.
[8] [1997] AC 749.
[9] [1998] 1 WLR 896.

would regard as relevant.[10] Relevance of the extrinsic evidence to the objective setting of the contract is the expressed criterion. It is, however, rare for a judge to decide that a text means something that it could not mean in ordinary or technical language. On the rare occasions when a judge does this, he does it because he thinks there is an obvious mistake which has been made by the author of the text and that he has a duty to correct it.

The decision in the *Investors Compensation Scheme* case raised questions about two sacred cows of English law, namely that the court is not permitted to use evidence of (1) pre-contractual negotiations of the parties or (2) of their subsequent conduct in aid of the construction of written contracts even if the material throws light on the subjective intentions of the parties.[11] One view is that these rules follow from the principle that the task of the court is simply to establish the objectively ascertained contextual meaning of the language of the contract. The other view is that the restrictive rules are imposed as a matter of legal policy to achieve certainty. It is, however, important to note that the Vienna Convention on the Sale of Goods (1980),[12] the Unidroit Principles of International Commercial Contracts,[13] and the Principles of European Contract Law (1998)[14] in principle permit such evidence to be taken into account. No doubt this liberality is due to the subjective approach of civil law system. Possibly we are swimming against the tide. In England the rule about prior negotiations may for the moment be relatively safe. I am less confident about the life expectancy of the rule excluding subsequent conduct. Business people and, for that matter, ordinary people, simply do not understand a rule which excludes from consideration how the parties have in the course of performance interpreted their contract. The law must not be allowed to drift too far from intuitive reactions of justice of men and women of good sense: the rule about subsequent conduct may have to be re-examined. In any event, the strict application of these rules had to be qualified in practice. Pragmatically, it has been decided that if pre-contractual exchanges show that the parties attached an agreed meaning to ambiguous expressions, that may be admitted in aid of interpretation.[15] That is a substantial inroad on the restrictive rules. The courts have resorted to estoppel to temper the rigidity of the orthodox rule regarding the inadmissibility of subsequent conduct. Thus in *Vistafjord* the Court of Appeal held that a party may be precluded by an estoppel by convention from raising a contention contrary to a common assumption of fact or law (including the interpretation of a contract)

[10] [2002] 1 AC 251, at 269 (para 39).
[11] *Prenn v Simmons* [1971] 1 WLR 1381; *James Millar and Partners v Whitworth Street Estates (Manchester) Ltd* [1970] AC 583.
[12] Art 8(3).
[13] Art 4(3).
[14] Art 5.102.
[15] *The Karen Oltmann* [1976] 2 Lloyd's Rep 708. See McLauchlan (1997) 113 LQR 237.

on which the parties have acted.[16] In this way the reasonable expectations of parties are given some protection. A more radical approach to the two restrictive rules may become necessary. It may be that the differences between commercial and consumer transactions should be more clearly recognised: a hard-nosed attitude to admitting such evidence in commercial transactions may be right but in consumer transactions a more relaxed approach may be necessary.

That brings me to the implication of terms.[17] It is part of the interpretative process. In systems of law where there is a general duty of good faith in the performance of contracts the need to supplement the written contract by implied terms is less than in the common law system. The implication of terms fulfils an important function in promoting the reasonable expectations of parties. Two types of implication are relevant. First, there are terms implied in fact, ie from the contextual scene of the particular contract. Such implied terms fulfil the role of ad hoc gap fillers. Often the expectations of the parties would be defeated if a term were not implied, eg sometimes a contract simply will not work unless a particular duty to cooperate is implied. The law has evolved practical tests for such an implication, such as the test whether the term is necessary to give business efficacy to the contract or whether the conventional bystander, when faced with the problem, would immediately say 'yes, it is obvious that there must be such a term'. The legal test for the implication of a term is the standard of strict necessity. And it is right that it should be so since courts ought not to supplement a contract by an implication unless it is perfectly obvious that it is necessary to give effect to the reasonable expectations of parties. It is, however, a myth to regard such an implied term as based on an inference of the actual intention of the parties. The reasonable expectations of the parties in an objective sense are controlling: they sometimes demand that such terms be imputed to the parties. The second category is terms implied by law. This occurs when incidents are impliedly annexed to particular forms of contracts, eg contracts for building work, contracts of sale, hire etc. Such implied terms operate as default rules. By and large such implied terms have crystallised in statute or case law. But there is scope for further development in a rapidly changing world. This function of the court is essential in providing a reasonable and fair framework for contracting. After all, there are many incidents of contracts which the parties cannot always be expected to reproduce in writing.

Ending my discussion of contracts, the black letter approach to interpretation of contracts has given way to a more commercial approach. It eschews niceties of language and concentrates on a contextual approach and the structure and purpose of the transaction. It is to be welcomed.

[16] [1988] 2 Lloyd's Rep 343.
[17] Andrew Phang, 'Implied Terms, Business Efficacy and the Officious Bystander' [1988] JBL 1.

On the other hand, the problem of the consumer perspective has not been completely solved.

II. STATUTE LAW

In 1882 Pollock described the approach of English judges to statutes as follows: 'Parliament generally changes law for the worse, and ... the business of the judges is to keep the mischief of its interference within the narrowest bounds.'[18] This was an accurate description of the judicial mindset in Victorian times. This approach led to restrictive interpretation by literalist methods which sometimes blocked social progress. It remained the approach of English judges until some time after the Second World War. But the legal world has changed. Like Australian judges, English judges now apply purposive methods of construction of statutes.

Except in the rare case where a statute reveals a contrary intention, it is now settled that every statute must be interpreted as an 'always speaking statute'. There are at least two strands to this principle. The first is that courts must interpret and apply a statute of any vintage to the world as it exists today. That is the basis of the decision of the House of Lords in *R v Ireland,* a case where 'bodily harm' in a Victorian statute governing assaults was held to cover psychiatric injury.[19] Equally important is the second strand, namely that a statute must be interpreted in the light of the legal system as it exists today.[20] Thus the importance the law nowadays attach to free speech is relevant background to the interpretation of earlier statutes. The rationale of this principle is that a statute is usually intended to endure for a long time in a changing world. This principle does not apply to contracts. Arguably, however, there could be a similar development in respect of international standard form contracts with an intended long life.

It will be rare for a statute to have one obvious meaning which can be determined without taking into account the context of the legislation. One might say that a statutory provision that a notice must be lodged within 30 days requires no resort to contextual material. But even this proposition is not necessarily correct. The context may throw light on the relative plausibility of interpretations holding that days include every day of the week or only week days. While the text of the statute is of pre-eminent importance, it cannot be understood in a vacuum.

[18] *Essays on Jurisprudence and Ethics* (1882) 85.
[19] [1988] AC 147, 158D-G.
[20] Cross, *Statutory Interpretation,* 3rd edn (1995), 51–52; *McCartan Turkington Breen v Times Newspapers Ltd* [2001] 2 AC 277, 296A-F.

This was lucidly explained by Lord Blackburn in *River Wear Commissioners v Adamson* as follows:[21]

> … I shall … state, as precisely as I can, what I understand from the decided cases to be the principles on which the Courts of Law act in construing instruments in writing; and a statute is an instrument in writing. In all cases the object is to see what is the intention expressed by the words used. But, from the imperfection of language, it is impossible to know what that intention is without inquiring farther, and seeing what the circumstances were with reference to which the words were used, and what was the object, appearing from those circumstances, which the person using them had in view; for the meaning of words varies according to the circumstances with respect to which they were used.

Legislative language can only be understood against the backcloth of the world to which it relates. Sometimes judgments do not fully take into account the different levels of reasoning at which the context is relevant. As in the case of commercial contracts, and other legal texts, the context is relevant to what possible different meanings the language of the text may let in. But the context is again relevant when the judge comes to select among the possible interpretations the best one. It is therefore a fundamental misconception to say that the background to the statute may only be admitted in the event of an ambiguity. The interpretative process requires judges to make informed choices.

That brings me to the use of Hansard in aid of interpretation of statutes. As in Australia, English courts regularly use reports, which led to or preceded legislation, in aid of interpretation. It is part of the setting of a statute. Australian and English lawyers would agree that there is no reason why Hansard materials should not be used to identify the mischief against which the statute is aimed. It helps to explain the background of the statute. Far more troublesome is the use in aid of interpretation of statements of a government minister in the promotion of a bill as reflecting the desired intent of the government. Section 15A-B of the Acts Interpretation Act 1901, as inserted in 1984, permits the use of such material where the legislation is ambiguous or obscure or its literal meaning leads to an absurdity. In England *Pepper v Hart*[22] heralded a parallel development. Doubts about the reach of that decision have arisen in England. It may be of interest if I explained the reservations which are now emerging in England.

Pepper v Hart broke new ground by holding that in cases of ambiguity it is permissible to refer in aid of construction of statutes to statements of a promoter of the bill. The rationale of this principle was memorably stated by Lord Denning:

> Why should judges grope about in the dark searching for the meaning of an Act, when they can so easily switch on the light?

21 (1877) 2 App Cas 743, 763.
22 [1993] AC 593.

I have, however, come to the conclusion that while the actual decision in *Pepper v Hart* was correct, the broadly based observations in that case are contrary to constitutional principle.[23]

In the Westminster Parliament exchanges sometimes take place late at night in nearly empty chambers whilst places of liquid refreshment are open. Sometimes there is a party political debate with whips on. The questions are often difficult but political warfare sometimes leaves little time for reflection. These are not ideal conditions for the making of authoritative statements about the meaning of a clause in a Bill. Let me give you the flavour from an explanation by Lord Hayhoe, reported in Hansard of 27 March 1996. He said:

> I remember only too well my first intervention as a new Minister at the Treasury on the Finance Bill in the very early hours of the morning on a subject about which I knew absolutely nothing but on which I had a marvellously thick book of briefing from the Inland Revenue. I appropriately read out the response to some detailed points that had been made by one of the Opposition spokesmen who stood up afterwards to say how well I had dealt with the point he had raised and welcomed my first intervention in Finance Bill Committees. However, I discovered from my private office afterwards that I had read out the wrong reply to the amendment. Clearly, it made not the slightest bit of difference.

It is sometimes meaningful and appropriate for a judge to refer to the intention of Parliament in recognition of its supreme law-making power. It is also perfectly sensible to say that legislation as duly promulgated reflects the will of Parliament. But it is quite a different matter to ascribe to a composite and artificial body such as a legislature a state of mind deduced from exchanges in debates. The law can ascribe to legal persons, such as companies and state agencies, an intention to commit particular acts. Rules of attribution have been developed to suit the demands of particular contexts. But the argument that a legislature, operating through two chambers, may have an intention revealed by statements in debates is altogether more ambitious. Until *Pepper v Hart,* under the common law, there was in England no rule of attribution, or rule of recognition, which treated statements of ministers as acts of Parliament.

The intention under consideration is one targeted on the meaning of language contained in a clause in a Bill and employed in a ministerial statement. A Bill is a unique document. It speaks in compressed language. Parliament legislates by the use of general words. It is difficult to ascribe to members of Parliament an intention in respect of the meaning of a clause in

[23] I have drawn on my paper '*Pepper v Hart:* A Re-examination' (2001) 21 *Oxford Journal of Legal Studies* 59.

a complex Bill and how it interacts with a ministerial explanation. The ministerial explanation in *Pepper v Hart* was made in the House of Commons only. What is said in one House in debates is not formally or in reality known to the members of the other House. How can it then be said that the minister's statement represents the intention of Parliament, ie both Houses. The Appellate Committee took the view that opposing views expressed by a person other than the promoter can safely be disregarded whenever a statement by a promoter is admitted. This is also an assumption which seems inherently implausible in respect of the ebb and flow of Parliamentary debates. In truth a minister speaks for the government and not for Parliament. The statements of a minister are no more than indications of what the government would like the law to be. In any event, it is not discoverable from the printed record whether individual members of the legislature, let alone a plurality in each chamber, understood and accepted a ministerial explanation of the suggested meaning of the words. For many the spectre of the ever watchful whips will be enough. They may agree on only one thing, namely to vote yes. And they have no means of voting yes and registering at the same time disagreement with the explanation of the minister. Their silence is therefore equivocal. When one considers such realities of Parliamentary life the idea of determining from Hansard the true intention of Parliament on the meaning of a clause in a Bill, and an associated ministerial statement, looks more and more far-fetched. In *Black-Clawson*,[24] Lord Reid, speaking with enormous Parliamentary experience, said, 'We often say that we are looking for the intention of Parliament but that is not quite accurate. We are seeking the meaning of the words which Parliament use.' It would have been a fiction for the House to say in *Pepper v Hart* that as a matter of historical fact the explanation of the Financial Secretary reflected the intention of Parliament. Arguably the House may have had in mind in *Pepper v Hart* that an intention derivable from the Financial Secretary's statement ought to be *imputed* to Parliament. If that were the case, the reasoning would rest on a complete fiction. The only relevant intention of Parliament can be the intention of the composite and artificial body to enact the statute as printed.

There is a strong case for allowing a statute to be interpreted in favour of the citizen in accordance with a considered explanation given by a minister promoting the Bill. It is the argument that the executive ought not to get away with saying in a Parliamentary debate that the proposed legislation means one thing in order to ensure the passing of the legislation and then to argue in court that the legislation bears the opposite meaning. That is what happened in *Pepper v Hart*. Lord Bridge of Harwich said that the Financial Secretary 'assured' the House that it was not intended to impose the

[24] *Black-Clawson International Ltd v Papierwerke Waldhof-Aschaffenburg AG* [1975] AC 591.

relevant tax. He must have taken the view, as did other members of the majority, that the Revenue in imposing the tax were going back on an assurance to the House of Commons. That would have been an unfair and unacceptable result. If such a consequence prevailed it would tend to undermine confidence in the legal system.

Whether one calls it an estoppel, a legitimate expectation, a principle of fairness, or whatever, *Pepper v Hart* as decided on its facts can simply be viewed as a tempering of the traditional exclusionary rule in the interests of justice. On this basis the impact of the decision can be confined to the admission *against* the executive of categorical assurances given by ministers to Parliament. This may be a principled justification of *Pepper v Hart.* And it does not involve a search for the phantom of a Parliamentary intention.

Unfortunately, that is not how the reasoning of the House in *Pepper v Hart* was expressed. The House had before it a ministerial statement which it regarded as favouring the taxpayer. This framework dictated the shape of the arguments and the judgments. The converse case was not considered. What would the position have been if the statutory position had been truly ambiguous and the ministerial statement favoured the Revenue? *Ex hypothesi* the statement would have come from a minister promoting the Bill and would have been clear on the very question in issue. It would therefore have been a trump card. A judge who declined to give effect to it would, on the reasoning in *Pepper v Hart*, be thwarting the intention of Parliament. What then happens to the principle that if a taxation provision is reasonably capable of two alternative meanings, the courts will prefer the meaning more favourable to the subject? *Pepper v Hart* does not address this question.

The basis on which the exclusionary rule was relaxed ignores constitutional arguments of substance. Lord Bridge described the rule as 'a technical rule of construction'. And implicitly that is how the majority approached the matter. Surely, it was much more. It was a rule of constitutional importance which guaranteed that only Parliament, and not the executive ultimately legislates; and that the courts are obliged to interpret and apply what Parliament has enacted, and nothing more or less. To give the executive, which promotes a Bill, the right to put its own gloss on the Bill is a substantial inroad on a constitutional principle, shifting legislative power from Parliament to the executive. Given that the ministerial explanation is *ex hypothesi* clear on the very point of construction, *Pepper v Hart* treats qualifying ministerial policy statements as canonical. It treats them as a source of law. It is in constitutional terms a retrograde step: it enables the executive to make law. It is of fundamental importance to understand that the objection is not to the idea of a judge looking at Hansard. It is entirely acceptable for a judge to identify the mischief of a statute from Hansard. What is constitutionally wrong in the English system is to treat the intentions of the government as revealed in debates as reflecting the will of Parliament.

A matter not considered in *Pepper v Hart* is the likely impact of the relaxation of the exclusionary rule on executive practice. It was always predictable that the behaviour of ministers would alter in response to the change announced in *Pepper v Hart*. After all, why should ministers not take advantage of *Pepper v Hart* to explain the effect of the legislation in the way in which the government would like it to be understood? If this happens it must mark a constitutional shift of power from Parliament to ministers. The Parliamentary debates leading to the enactment of the Human Rights Act 1998 are revealing. When questioned about the effect of the omission to incorporate Article 13 of the European Convention on Human Rights the Lord Chancellor said: 'One always has in mind *Pepper v Hart* when one is asked questions of that kind. I shall reply as candidly as I may.'[25] This makes my point: executive practice is bound to be influenced by *Pepper v Hart*. There is a real incentive for government to use this strategy to get on the record *Pepper v Hart* statements when it is reluctant to spell out its precise intentions on the face of the bill.

In *Pepper v Hart* the House of Lords failed to consider important constitutional questions. There are, however, those who believe that the relaxation of the exclusionary rule was the ultimate vindication of purposive construction. And purposive construction is like mother's milk and apple pie: who can argue against it? The reasoning in *Pepper v Hart* sought to build on the fact that official reports and white papers are admissible for the purpose of identifying the mischief to be corrected. Such reports are always admissible for what logical value they have. But the constitutional objections do not apply to such reports. They are part of the contextual scene against which Parliament legislates. In any event, to present the *Pepper v Hart* issue as depending on whether one adopts a literal or purposive approach to construction is wide off the mark. By the time *Pepper v Hart* was decided nobody supported literal methods of construction. The suggested antithesis misses the point of the fundamental and constitutional nature of the objections. The objections are not simply that a minister's view of a clause is irrelevant but that it is in principle wrong to treat it as a trump card or even relevant in the interpretative process.

What are the chances of *Pepper v Hart* being reversed? Being a decision that marks a shift of power from Parliament to the executive, the prospect of any government initiating legislation to reverse it must be slight. It is, however, possible that *Pepper v Hart* may be confined by judicial decision to the use of Hansard *against* the executive when it goes back on an assurance given to Parliament. This would not require the overruling of *Pepper v Hart*. It would simply confine its legal force to the material circumstances of that case. In England this question will not go away. In two recent

[25] Hansard, HL (18 November 1997) Col 475.

decisions in the House of Lords there have been dicta raising these questions.[26] The debate continues.

III. CONSTITUTIONAL MEASURES: FUNDAMENTAL RIGHTS

Constitutional adjudication affecting fundamental rights contained in a bill of rights requires a broader approach than is applicable to commercial contracts and statutes. It requires what Lord Wilberforce in *Foster* described as 'a generous interpretation avoiding what has been called 'the austerity of tabulated legalism', suitable to give to individuals the full measure of fundamental rights and freedoms'.[27]

Bills of rights have proved themselves in the common law world, influencing the common law and being influenced by it. The interpretative techniques adopted in common law countries vary. The Canadian Charter of Rights and Freedoms requires judges to interpret an impugned law in a way that conforms to the Charter. If it cannot be reconciled the court declares the inconsistency and the law is pro tanto void. The New Zealand Bill of Rights 1990 is a weaker model. While the court must strive to reach an interpretation compatible with the Bill of Rights, there is no express power for the court to go further. On the other hand, the New Zealand Court of Appeal has strengthened the regime by holding that there is an implied power make a declaration of inconsistency.[28] With the advantage of these earlier models the South African Constitution, and its Bills of Rights, was carefully crafted to entrench human rights strongly. Unlike Canada and New Zealand there is a Constitutional Court to adjudicate on constitutional issues. It has the power to declare legislation unconstitutional. The United Kingdom Bill of Rights is a relative newcomer in the field. The Human Rights Act 1998, which incorporates the European Convention on Human Rights into our law, came into force in October 2000. There is a strong interpretative obligation on the court to interpret legislation so as to be compatible with the Convention. If it is impossible to do so, the court must make a declaration of incompatibility. Parliamentary supremacy is respected. The expectation is, however, that Parliament will consider the law on an early occasion and amend it.

In advance of the Human Rights Act 1998 coming into operation there was much hysteria. Newspapers described the Act as a recipe for chaos. They feared that traffic would be brought to a halt; that serious crime

[26] *Regina v Secretary of State for the Environment, Transport and Regions, ex parte Spath Holme* [2001] 2 AC 349, at 407 (*per* Lord Hope of Craighead); *Robinson v Secretary of State for Northern Ireland and Others* [2002] UKHL 32, *per* Lord Hoffmann and Lord Millett.
[27] *Ministry of Home Affairs v Fisher* [1980] AC 319, at 328.
[28] *Moohen v Film & Literature Board of Review* [2000] 2 NZLR 9; *Poumaka* [2000] 2 NZLR 37.

would go unpunished; and that the prison gates would be left permanently open. The forecast was that it would rain sulphur and brimstone. The premise of this hysteria was that the courts would accede to every impractical and implausible claim, ignoring the balance inherent in the Convention between individual rights and conditions of stability and order required in a liberal democracy. This ignored the fact that the direct application of the Convention has caused no such chaos in other European democracies. Nevertheless, Lord McCluskey, a Scottish judge, joined in by saying that the Human Rights Act 'would provide a field-day for crackpots, a pain in the neck for judges and legislators, and a goldmine for lawyers'. Unsurprisingly, the judge was held to be disqualified from sitting in a human rights case.[29] The predicted legal revolution has not taken place. Instead there has been a subtle process of weaving human rights law into United Kingdom law. The Act has bedded down in a sensible and satisfactory way. Only a small percentage of challenges have succeeded. But it has afforded an opportunity for the courts to examine critically but constructively a few murkier areas of English law. It has strengthened our democracy.

It is undoubtedly the case that human rights are protected at many levels in the Australian system.[30] But unlike Western European countries and major Commonwealth countries, Australia has no express bill of rights. For Australian courts, fulfilling their constitutional duties of standing between the people and the executive, and protecting fundamental rights, this is a disadvantage.

Much can, however, be done through the common law. In England, the courts have recognised certain fundamental rights as constitutional. The courts protect as constitutional the right of participation in the democratic process, equality of treatment, freedom of expression, religious freedom and the right of unimpeded access to the courts. Even before the incorporation of the European Convention on Human Rights into English law the courts held that everybody has an absolute constitutional right to a fair trial which if breached must lead to the setting aside of the conviction.[31] What is the significance of classifying a right as constitutional? It is meaningful. It is a powerful indication that added value is attached to the protection of the right. It strengthens the normative force of such rights.[32] It virtually rules out arguments that such rights can be impliedly repealed by subsequent legislation.[33] Generally only an express repeal will suffice. The constitutionality of a right is also important in regard to remedies. The duty of the court is to

[29] *Hoekstra and Others v HM Advocate* [2001] 1 AC 216.
[30] George Williams, *Human Rights Under the Australian Constitution,* (1999), *passim.*
[31] *R v Brown (Winston)* [1998] AC 367; *R v Bentley* [2001] 1 Cr App Rep 307. Now the absolute guarantee of a fair trial is governed by Art 6.1 of the European Convention: the relevant case law is reviewed in *Mills v HM Advocate and Another* [2002] 3 WLR 1597.
[32] *Mohammed v The State* [1999] 2 AC 111.
[33] *Thoburn and Others v Sunderford City Council and Others* [2003] QB 151.

vindicate the breach of a constitutional right, depending on its nature, by an appropriate remedy.

There is another important common law development. Parliament does not legislate on a blank sheet. In the case of Britain it legislates for a European liberal democracy. This gives rise to what Sir Rupert Cross described as a presumption of general application which operates as a constitutional principle.[34] General words in a statute should not be allowed to abrogate fundamental rights. This principle has a considerable common law pedigree but in practice judges often failed to observe it. In 1998 in *Pierson*[35] in separate judgments Lord Browne-Wilkinson and I tried to bring together the rich strands of authority in support of this principle. At that time our views did not attract the support of a majority. Two years later in *Simms*[36] the House of Lords unambiguously reaffirmed the principle. Lord Hoffmann explained the rationale of the principle:

> Parliamentary sovereignty means that Parliament can, if it chooses, legislate contrary to fundamental principles of human rights .... The constraints upon its exercise by Parliament are ultimately political, not legal. But the principle of legality means that Parliament must squarely confront what it is doing and accept the political cost. Fundamental rights cannot be overridden by general or ambiguous words. This is because there is too great a risk that the full implications of their unqualified meaning may have passed unnoticed in the democratic process. In the absence of express language or necessary implication to the contrary, the courts therefore presume that even the most general words were intended to be subject to the basic rights of the individual.

This principle is now firmly entrenched in English law.[37] If it is applied by your courts, it will strongly reinforce the protection of fundamental rights in Australia.

IV. CONCLUSION

I end by saying that in the interpretative process the judiciary owes allegiance to nothing except the constitutional duty of reaching through reasoned debate the best attainable judgments in accordance with justice and law. This is their role in the democratic governance of our countries.[38]

[34] *Statutory Interpretation*, 3rd edn, at 166.
[35] [1998] AC 539, 575D.
[36] *R v Secretary for the Home Department ex parte Simms* [2000] 2 AC 131.
[37] *R v Special Commissioner and Another, ex parte Morgan Grenfell & Co Ltd* [2003] 1 AC 563.
[38] I am indebted to Lydia Clapinska, a Judicial Assistant in the House of Lords, now based in Sydney, and to Karen Steyn, for suggestions about my lecture.

6

The Interpretation of Contracts: Lord Hoffmann's Re-Statement

EWAN McKENDRICK

THIS CHAPTER AIMS to consider the impact of Lord Hoffmann's re-statement of the law relating to the interpretation of contracts in *Investors Compensation Scheme Ltd v West Bromwich Building Society*.[1] It is based on a survey of some 180 English[2] reported and unreported cases in which Lord Hoffmann's speech was considered.[3] The number of citations may occasion some surprise. *Investors Compensation Scheme* has not received a great deal of attention in textbooks on the law of contract,[4] in the law reviews[5] or, I suspect, in undergraduate legal education. But it has received a great deal of attention in legal practice. The reason for this is not hard to seek. Many contract disputes turn on the proper interpretation of the contract between the parties. Commercial parties have their own standard terms and conditions of business and the meaning of these standard terms (often referred to as 'boilerplate clauses') is a matter of great significance to the parties.

[1] [1998] 1 WLR 896, 912–913. Lord Hoffmann's re-statement is not the only one that has been made in recent years. There have been others: by Colman J in *Zeus Tradition Marine Ltd v Bell (The 'Zeus V')* [1999] 1 Lloyd's Rep 703, 706–707, by Judge Anthony Thornton Q C in *Demolition Services Ltd v Castle Vale Housing Action Trust*, (1999) 79 Con LR 55, by Clarke L J in *Martinez v Ellesse International SpA* Unreported, 30 March 1999, Court of Appeal and by Lightman J in *Household Global Funding Inc v British Gas Trading Ltd,* unreported, 29 June 2001, Chancery Division. But none of these re-statements has acquired the authority of Lord Hoffmann's re-statement.

[2] Scottish cases and decisions from other jurisdictions have not been included in this survey. There have, in fact, been a number of Scottish cases in which consideration has been given to Lord Hoffmann's re-statement.

[3] I am grateful to Felicity Maher who undertook the burden of collecting and summarising the case-law. The survey is based on a LEXIS search and was completed on 14 October 2002.

[4] See for example, Treitel *The Law of Contract* 10th edn, (1999) pp 812–813 where the case is mentioned in a couple of the footnotes.

[5] For exceptions see DW McLauchlan 'The New Law of Contract Interpretation' (2000) 19 *New Zealand Universities Law Review* 147 and G McMeel 'The Rise of Commercial Construction in Contract Law' [1998] *LMCLQ* 382.

This chapter is no more than an attempt to draw together the various strands that are identifiable in the case law. Detailed consideration of the various issues raised must await another day. At this stage all that can be done is to note some broad themes that emerge from the case law and then to explore, albeit briefly, the principal issues that have arisen in litigation post-*Investors Compensation Scheme*. Three broad themes can be identified. The first relates to the extent of the change introduced by Lord Hoffmann. Is his speech a revolutionary one or does it simply reflect the fact that, over the last 30 or so years, fundamental changes have taken place in the way in which the courts interpret contracts? Further, what impact does his judgment have on previous case law? In *NLA Group Ltd v Bowers*[6] Timothy Walker J stated that 'Lord Hoffmann was simply overruling old and outdated cases by reference to an approach on construction which has been followed in the Commercial Court for many years.'[7] But which 'old and outdated cases' have been overruled? As we shall see, this is an issue that has given rise to some uncertainty in connection with the application of Lord Hoffmann's re-statement to the construction of exclusion clauses. The second theme relates to the range of materials to which a court can have regard when seeking to ascertain the meaning of the words in a contract. The range of materials to which the courts can have regard has clearly increased in recent years and, subject to two exceptions, it can be said that attention has shifted from the question whether or not the evidence is **admissible** to the question of the **weight** that is to be given to that evidence by the judge. Some judges have expressed concern about this development. Their concerns appear to be based on two grounds. The first is one of cost. The admission of more evidence is likely to increase the length and hence the cost of litigation. These additional costs may not be outweighed by any benefits obtained as a result of permitting the courts to have regard to a wider range of materials. The second objection relates to the uncertainty created by this extension, particularly as a consequence of Lord Hoffmann's invocation of the phrase 'matrix of fact' which was borrowed from the speech of Lord Wilberforce in *Prenn v Simmonds*.[8] Some judges, most notably Sir Christopher Staughton,[9] have objected that the phrase is too uncertain and that this uncertainty can be exploited by counsel in order to adduce evidence, the relevance of which is, at best, doubtful. The third theme relates to the balance struck by the courts between the need to pay careful attention to the language used by the parties and the concern of the court to give effect to the perceived commercial purpose of

[6] [1999] 1 Lloyd's Rep 109.
[7] *Ibid*, at p 112.
[8] [1971] 1 WLR 1381.
[9] For his extra-judicial views see Sir Christopher Staughton, 'How do the Courts Interpret Commercial Contracts?' [1999] *Cambridge Law Journal* 303, and for similar views expressed in a judicial capacity see *Scottish Power plc v Britoil (Exploration) Ltd, The Times*, 2 December 1997.

the clause or contract in issue. Do the courts have primary regard to the language in which the parties have chosen to express their agreement or is primacy to be given to the commercial purpose or to the adoption of a commercially sensible construction? This is an issue which has produced some judicial difference of opinion post-*Investors Compensation Scheme*.

This essay proceeds in three stages. The first part simply sets out Lord Hoffmann's speech in *Investors Compensation Scheme*. The second part deals with judicial reaction to Lord Hoffmann's re-statement; in particular, it focuses attention on judicial criticisms of his approach. The third part considers the scope of the principles laid down by Lord Hoffmann and addresses a number of issues that have arisen in relation to their scope. It is not possible within the scope of this essay to deal with these issues in any detail. All that can be done is to sketch out some of the difficulties that the courts have experienced in applying Lord Hoffmann's re-statement.

I. LORD HOFFMANN'S RE-STATEMENT

Investors Compensation Scheme Ltd v West Bromwich Building Society[10] was a case concerned with the construction of a claim form. The investors submitted a claim to the Investors Compensation Scheme Ltd ('ICS') for compensation in respect of negligent advice given by brokers who had sold them home income plans in contravention of the provisions of the Financial Services Act 1986. The plans proved to be a disastrous investment and rendered the investors' homes vulnerable to re-possession by the West Bromwich Building Society. The claim form was drafted by the ICS and it required the investors to assign to ICS all their rights arising out of the sale to them of the home income plans. Excluded from the assignment was 'any claim (whether sounding in rescission for undue influence or otherwise) that you [the investor] may have against [the building society] in which you claim an abatement of sums which you would otherwise have to pay to that Society in respect of sums borrowed by you from that Society in connection with the transaction and dealings giving rise to the claim.' The investors then brought proceedings against the West Bromwich Building Society seeking damages for negligence, misrepresentation and rescission of their mortgages. ICS also instituted proceedings against the building society in which it asserted that the investors had assigned all their claims to ICS with the exception of the investors' claim for rescission. The claims brought by the investors and ICS against the building society therefore overlapped. The issue of interpretation which arose between the parties related to the construction of the claim form, in particular the scope of the provisions relating to the assignment of the investors' rights against third parties.

[10] [1998] 1 WLR 896.

Lord Hoffmann prefaced his remarks on the particular issue of construction with the following general observations:

> I think I should preface my explanation of my reasons with some general remarks about the principles by which contractual documents are nowadays construed. I do not think that the fundamental change which has overtaken this branch of the law, particularly as a result of the speeches of Lord Wilberforce in *Prenn v Simmonds* [1971] 1 WLR 1381 at 1384–1386 and *Reardon Smith Line Ltd v Hansen-Tangen, Hansen-Tangen v Sanko Steamship Co* [1976] 1 WLR 989, is always sufficiently appreciated. The result has been, subject to one important exception, to assimilate the way in which such documents are interpreted by judges to the common sense principles by which any serious utterance would be interpreted in ordinary life. Almost all the old intellectual baggage of 'legal' interpretation has been discarded. The principles may be summarised as follows.
>
> (1) Interpretation is the ascertainment of the meaning which the document would convey to a reasonable person having all the background knowledge which would reasonably have been available to the parties in the situation in which they were at the time of the contract.
>
> (2) The background was famously referred to by Lord Wilberforce as the 'matrix of fact', but this phrase is, if anything, an understated description of what the background may include. Subject to the requirement that it should have been reasonably available to the parties and to the exception to be mentioned next, it includes absolutely anything which would have affected the way in which the language of the document would have been understood by a reasonable man.
>
> (3) The law excludes from the admissible background the previous negotiations of the parties and their declarations of subjective intent. They are admissible only in an action for rectification. The law makes this distinction for reasons of practical policy and, in this respect only, legal interpretation differs from the way we would interpret utterances in ordinary life. The boundaries of this exception are in some respects unclear. But this is not the occasion on which to explore them.
>
> (4) The meaning which a document (or any other utterance) would convey to a reasonable man is not the same thing as the meaning of its words. The meaning of words is a matter of dictionaries and grammars; the meaning of the document is what the parties using those words against the relevant background would reasonably have been understood to mean. The background may not merely enable the reasonable man to choose between the possible meanings of words which are ambiguous but even (as occasionally happens in ordinary life) to conclude that the parties must, for whatever reason, have used the wrong words or syntax (see *Mannai Investment Co Ltd v Eagle Star Life Assurance Co Ltd* [1997] 2 WLR 945).
>
> (5) The 'rule' that words should be given their 'natural and ordinary meaning' reflects the commonsense proposition that we do not easily accept that

> people have made linguistic mistakes, particularly in formal documents. On the other hand, if one would nevertheless conclude from the background that something must have gone wrong with the language, the law does not require judges to attribute to the parties an intention which they plainly could not have had. Lord Diplock made this point more vigorously when he said in *Antaios Cia Naviera SA v Salen Rederierna AB, The Antaios* [1985] AC 191 at 201:
>
> ... if detailed semantic and syntactical analysis of words in a commercial contract is going to lead to a conclusion that flouts business common sense, it must be made to yield to business common sense.[11]

Applying these principles to the facts of the case Lord Hoffmann concluded that all claims for damages and compensation had been validly assigned to ICS (so that ICS and not the investors could bring a claim for damages) but that the investors had retained the right to claim rescission of the mortgages. Thus he concluded that the phrase 'Any claim (whether sounding in rescission for undue influence or otherwise)' was actually used by the parties to mean 'Any claim sounding in rescission (whether for undue influence or otherwise)'. In reaching this conclusion Lord Hoffmann accepted that he had not given the words their natural and ordinary meaning but the reason for this was that this was not the way in which the parties had used the words. The court was therefore 'engaged in choosing between competing unnatural meanings.'[12]

Two preliminary points should be noted about Lord Hoffmann's 'general remarks'. The first is that they do not purport to be revolutionary. Rather, they attempt to build on the principles established by Lord Wilberforce in *Prenn v Simmonds* and *Reardon Smith Line Ltd v Hansen-Tangen.* While it is true that Lord Hoffmann notes that there has been a fundamental change in the applicable legal principles, he attributes that change to Lord Wilberforce and not to himself. Secondly, the immediate context of his re-statement of the principles by which contracts are to be interpreted was a case in which the parties had not used words in their natural and ordinary meaning.

II. JUDICIAL REACTION TO LORD HOFFMANN'S RE-STATEMENT

Initial judicial response to Lord Hoffmann's re-statement was distinctly cool. This can be seen in *Investors Compensation Scheme* itself, where Lord Lloyd dissented in robust terms. He stated that he knew of 'no principle of construction' which enabled a court 'to take words from within the brackets, where they are clearly intended to underline the width of "any claim",

11 [1998] 1 WLR 896, 912–913.
12 *Ibid*, at p 914.

and place them outside the brackets where they have the exact opposite effect.'[13] What the parties meant was to be derived from the words they had used. He continued:

> purposive interpretation of a contract is a useful tool where the purpose can be identified with reasonable certainty. But creative interpretation is another thing altogether. The one must not be allowed to shade into the other.[14]

The decision of the House of Lords in *Investors Compensation Scheme* was given on 19 June 1997. Judicial response followed swiftly. The first case was the decision of the Court of Appeal in *National Bank of Sharjah v Dellborg*.[15] In fact, *Investors Compensation Scheme* was cited to the court only after the conclusion of argument but it is nevertheless mentioned in the judgments of Saville and Judge LJJ delivered on 9 July 1997. Saville L J only made a passing reference to *Investors Compensation Scheme* at the end of his judgment when he stated that nothing in the case caused him to alter the conclusion that he had reached on the facts of the case. He nevertheless launched an attack on the state of the law relating to the interpretation of contracts. He concluded that 'what appears to be the present law does no service to the international or domestic commercial community.' In particular, he stated that 'where the words used have an unambiguous and sensible meaning as a matter of ordinary language, I see serious objections in an approach which would permit the surrounding circumstances to alter that meaning.' He identified two principal objections in the following terms:

> Firstly, such an approach would seem to entail that even where the words that the parties have chosen to use have only one meaning; and that meaning (bearing in mind the aim or purpose of the agreement) is not self-evidently nonsensical, the court will not be allowed to adopt that meaning without an examination of the surrounding circumstances, which could involve discovery, interrogatories, cross-examination and the like; for a party seeking to challenge that meaning could assert with great force that until the circumstances are fully examined, it is impossible to decide whether or not they should override the plain words of the agreement. This would do nothing but add to the costs and delays of litigation and indeed of arbitration, much of which is concerned with interpreting agreements.
>
> Secondly, the position of third parties (which would include assignees of contractual rights) does not seem to have been considered at all. They are unlikely in the nature of things to be aware of the surrounding circumstances. Where the words of the agreement have only one meaning, and that meaning is not self evidently nonsensical, is the third party justified in taking that to be the agreement that was made, or unable to rely on the words used without examining

[13] *Ibid*, at p 904.
[14] *Ibid*.
[15] Unreported, Court of Appeal, 9 July 1997.

> (which it is likely to be difficult or impossible for third parties to do) all the surrounding circumstances? If the former is the case, the law would have to treat the agreement as meaning one thing to the parties and another to third parties, hardly a satisfactory state of affairs. If the latter is the case, then unless third parties can discover all the surrounding circumstances and are satisfied that they make no difference, they cannot safely proceed to act on the basis of what the agreement actually says. This again would seem to be highly unsatisfactory.

Judge LJ gave greater consideration to *Investors Compensation Scheme*. After setting out excerpts from Lord Hoffmann's fourth and fifth principles he stated:

> It therefore appears that if one party to the written contract wishes to dispute the obvious meaning of the words actually used in the document, as I understand it, he is entitled to do so by canvassing all the permitted matters of background—'absolutely anything'—which will serve to undermine the conclusion that the contract means what it apparently says. If this is the correct conclusion then I am bound to say expressly that I share the concerns expressed by Saville L J in his judgment.

Further judicial criticism of Lord Hoffmann's speech followed when, on 18 November 1997, the Court of Appeal gave judgment in the case of *Scottish Power plc v Britoil (Exploration) Ltd*.[16] Staughton L J was particularly critical of the width of Lord Hoffmann's second principle and his use of the phrase 'factual matrix'. He stated:

> it is often difficult for a judge to restrain the enthusiasm of counsel for producing a great deal of evidence under the heading of matrix, which on examination is found to contribute little or nothing to the true understanding of the parties' contract. All, or almost all, judges are now concerned about the huge cost of litigation. I have to say that such a wide definition of surrounding circumstances, background or matrix seems likely to increase the cost, to no very obvious advantage.

There are at least four aspects to these criticisms of Lord Hoffmann's re-statement. The first is the claim that it will add to the already extremely high cost of litigation. The second and related point is that it will generate uncertainty. Uncertainty is said to arise at a number of points, namely in relation to the scope of the 'factual matrix,' the possibility that the courts can resort to the surrounding circumstances even in the case where the words used clearly have a natural and ordinary meaning and in terms of the impact of the decision on older authorities. The third is that it has the potential to cause unfairness, particularly in its application to third parties

[16] *The Times,* 2 December 1997.

who may have no knowledge of the commercial purpose of the transaction and who rely solely on the words used by the parties. Finally Lord Hoffmann's re-statement may tempt judges to engage in the re-writing of contracts under the guise of the adoption of a commercial construction. These criticisms have all surfaced in the many decisions in which consideration has been given to the scope of Lord Hoffmann's speech and it is to these cases that we now turn.

III. THE SCOPE OF LORD HOFFMANN'S PRINCIPLES

As at 14 October 2002 the speech of Lord Hoffmann in *Investors Compensation Scheme* had been cited in some 180 decisions in the English courts.[17] Given the number of citations, it is hardly surprising that Waller L J in *Otelcilik AS v Hotel Management Corporation of America Ltd*[18] referred to Lord Hoffmann's 'now famous' statement of principles. In 25 of these cases *Investors Compensation Scheme* was cited by counsel but not by the judge or judges. These cases can therefore be laid to one side. In a further 30 cases the reference by the judge or judges to Lord Hoffmann's speech was extremely brief, often taking the form of a citation of the first or the second of Lord Hoffmann's principles. But in a number of cases the discussion was more extended. This section of the paper will focus on these cases. A number of issues will be canvassed, namely (i) the concern over costs, (ii) the abandonment of the 'old intellectual baggage of legal interpretation,' (iii) the related problem of the role of precedent in interpretation cases, (iv) the existence or otherwise of a revolution in the law relating to the interpretation of contracts, (v) the admissibility of pre-contractual negotiations, (vi) the admissibility of conduct subsequent to the making of the contract, (vii) the circumstances in which the courts are entitled to depart from the natural and ordinary meaning of the words used by the parties, (viii) the relationship between Lord Hoffmann's principles and the more general move, noted in particular in a number of Lord Steyn's speeches,[19] towards the adoption of a purposive approach towards interpretation, (ix) the definition of the 'matrix of fact' and (x) the distinction between 'commercial construction' and 're-writing' the contract for the parties. I shall comment briefly on each of these issues in turn.

[17] This figure is based on a Lexis survey and, in all probability, there are other cases in which Lord Hoffmann's speech was considered but which have not been retrieved by the Lexis search.

[18] Unreported, Court of Appeal, 25 May 1999.

[19] See, in particular, *Deutsche Genossenschaftsbank v Burnhope* [1995] 1 WLR 1580, 1589; *Lord Napier and Ettrick v R F Kershaw Ltd* [1999] 1 WLR 756, 763; *Mannai Investment Co Ltd v Eagle Star Life Assurance Co Ltd* [1997] AC 749, 770 and *Total Gas Marketing Ltd v Arco British Ltd* [1998] 2 Lloyd's Rep 209, 221.

A. The Problem of Cost

Cost was a problem expressly identified by both Saville L J in *National Bank of Sharjah v Dellborg* and by Staughton L J in *Scottish Power v Britoil*. This issue was addressed by Timothy Walker J in *NLA Group Ltd v Bowers,*[20] a case decided a year after the House of Lords gave judgment in *Investors Compensation Scheme*. The issue in the case was whether or not the plaintiff was entitled to receive a profit share commission under a Lloyd's slip policy. The issue of construction was a short one and it was tried by way of a preliminary issue. Notwithstanding the fact that the preliminary issue concerned a 'short point of construction of a few words'[21] counsel for the plaintiff 'originally had in mind to adduce the factual evidence of no fewer than five witnesses, with subpoenas served on two further witnesses'[22] together with an expert's report. Timothy Walker J decided the point of construction against the plaintiff and stated that 'the decision to attempt to adduce the factual evidence which the plaintiff attempted to adduce in this case was wholly unreasonable.'[23] Further, he concluded that there was a 'basis for ordering indemnity costs'[24] although he in fact ordered that the plaintiff pay half of the defendant's costs at the hearing before him on an indemnity basis.

The issue of cost has not arisen in such clear terms in subsequent cases, perhaps because counsel and the parties have taken note of the jurisdiction of the court to award costs on an indemnity basis. But this is not to say that the effect of the decision has not been to increase the cost of litigation. There appear to be very few cases in which the outcome was affected by reference to evidence of the surrounding circumstances or the wider context. In the vast majority of cases judges use the evidence of surrounding circumstances to confirm the decision they state they have already reached on the wording of the contract. To the extent that this evidence plays a confirmatory rule it could be said that the cost of adducing it, in terms of the time of the court and the hourly rate of lawyers, outweighs the benefits which it produces. It may also be the case that one consequence of Lord Hoffmann's approach has been to make it more difficult for the parties to obtain a preliminary ruling on a point of construction because the court is not confined to the words used by the parties but can have regard to a wider range of factors. The wider the range of materials to which the courts can have regard, the more likely it is that the parties will have to proceed directly to trial given that they will, in all probability, want to cross-examine witnesses etc on the content of their witness statements.

[20] [1999] 1 Lloyd's Rep 109.
[21] *Ibid*, at p 111.
[22] *Ibid*.
[23] *Ibid*, at p 113.
[24] *Ibid*.

B. Getting Rid of the Old Baggage

At first sight Lord Hoffmann's statement to the effect that 'almost all the old intellectual baggage of "legal" interpretation has been discarded'[25] does not appear to be particularly significant. But first appearances can deceive. The true significance of this sentence was explained by Lord Hoffmann in *Bank of Credit and Commerce International SA v Ali*[26] when, after referring to the judgments of Lord Denning in *George Mitchell (Chesterhall) Ltd v Finney Lock Seeds Ltd*[27] and Lord Wilberforce in *Photo Production Ltd v Securicor Transport Ltd*[28] he concluded that

> the disappearance of artificial rules for the construction of exemption clauses seems to me in accordance with the general trend in matters of construction which has been to try to assimilate judicial techniques of construction to those which would be used by a reasonable speaker of the language in the interpretation of any serious utterance in ordinary life.[29]

The cases post-*Investors' Compensation Scheme* do provide some support for a more relaxed approach to the interpretation of exclusion clauses and judges have cited *Investors Compensation Scheme* when adopting a more clement approach.[30] But it may be too soon to write the obituary for the old rules applicable to the interpretation of exclusion clauses. Some of the cases in which these rules were enunciated have been approved by appellate courts and cannot easily be swept aside. This is particularly so in relation to the construction of exclusion clauses that purport to exclude liability for negligence. The source of these rules can be traced back to the judgment of Lord Morton in *Canada Steamship Lines Ltd v The King*[31] and these rules have subsequently been applied by both the House of Lords[32] and the Court of Appeal.[33] Have these rules now 'disappeared'? It is unlikely that a single sentence expressed in general terms in *Investors Compensation Scheme,* combined with Lord Hoffmann's elaboration in *BCCI v Ali*, will suffice to overrule the *Canada Steamship* line of authority. Yet the relationship between these cases is an uneasy one. This can be demonstrated by

[25] [1998] 1 WLR 896, 912.
[26] [2001] UKHL 8, [2002] 1 AC 251.
[27] [1983] QB 284, 296–297.
[28] [1980] AC 827.
[29] [2001] UKHL 8, [2002] 1 AC 251, para [62].
[30] See, for example, *National Westminster Bank v Utrecht-America Finance* Co [2001] EWCA Civ 658; [2001] 3 All ER 733; *Pegler Ltd v Wang (UK) Ltd* [2000] BLR 218; *Stent Foundations Ltd v M J Gleeson Group Ltd* [2000] BLR 134, and *British Fermentation Products Ltd v Compair Reavell Ltd* [1999] BLR 352.
[31] [1952] AC 192.
[32] *Smith v UBM Chrysler (Scotland) Ltd* 1978 SC (HL) 1.
[33] See, for example, *The Raphael* [1982] 2 Lloyd's Rep 42 and *EE Caledonia Ltd v Orbit Valve Co Europe* [1994] 1 WLR 1515.

reference to the decision of the Court of Appeal in *National Westminster Bank v Utrecht-America Finance Company*[34] and the later decision of the House of Lords in *HIH Casualty and General Insurance Ltd v Chase Manhattan Bank*.[35] The issue before the Court of Appeal in the *National Westminster Bank* case was whether or not the *Canada Steamship* rules applied to a clause (clause 8.2(d)) that, in essence, purported to exclude liability for both negligent and fraudulent non-disclosure. Clarke L J referred to the *Canada Steamship* rules and stated that Lord Morton's rules 'are essentially rules or principles of construction.'[36] He stated that they remained 'valuable tools' in the construction of the contract but then attempted to link them with the rules of construction set out by Lord Hoffmann in *Investors Compensation Scheme* by stating that 'like any clause, cl 8.2(d) must be construed in its context and in the context of the contract as a whole having regard to its factual matrix or surrounding circumstances.'[37] Clarke L J concluded that clause 8.2(d) was 'a far cry from the kind of clause which the courts have had in mind in the kind of case in which they have applied Lord Morton's tests.'[38] The background to clause 8.2(d) was that the parties were two large banks, dealing at arm's length and advised by skilled commercial lawyers. The terms of the contract demonstrated that they had agreed that neither party was under a duty to disclose information to the other. He concluded, applying Lord Hoffmann's approach, that in these circumstances there was 'no room for the application of Lord Morton's tests'[39] and that effect should be given to the 'clear and unambiguous' terms of the clause which was that 'neither bank can be liable for breach of a duty to disclose, whether the failure to disclose is categorised by a foreign law as negligent, fraudulent or otherwise.'[40] However it should be noted that the case was concerned with liability for non-disclosure and that the Court of Appeal expressly recognised that the position would have been different had the clause purported to exclude liability for misrepresentation.[41]

There is a tension in the judgment of Clarke L J. On the one hand, he acknowledged that Lord Morton's tests remained 'valuable tools' but on the other hand he concluded that there was 'no room' for their application

[34] [2001] EWCA Civ 658; [2001] 3 All ER 733.
[35] [2003] UKHL 6.
[36] [2001] EWCA Civ 658; [2001] 3 All ER 733, para [47].
[37] Para [48].
[38] Para [49].
[39] Para [52].
[40] Para [54]. A different result was, however, reached by the House of Lords in *HIH Casualty and General Insurance Ltd v Chase Manhattan Bank* [2003] UKHL 6 where it was held that a similarly worded clause was not effective, as a matter of construction, to take away the right of the insurer to rescind the contract of insurance on the ground of fraudulent non-disclosure on the part of the agent of the insured.
[41] Paras [54] and [59].

to the clause in issue between the parties. A similar ambivalence can be found in the speeches of their Lordships in *HIH Casualty and General Insurance Ltd v Chase Manhattan Bank*.[42] Thus Lord Bingham stated that the tests laid down by Lord Morton retain 'general authority'[43] but at the same time he emphasised that Lord Morton was 'not laying down a code.'[44] The significance of Lord Morton's tests would appear to be on the decline in at least two respects. First, their Lordships in *HIH Casualty and General Insurance Ltd v Chase Manhattan Bank* emphasised that the paramount task of the court is to give effect to the intention of the parties:[45] the guidelines laid down by Lord Morton are tools to be used by the courts and they are not their masters. To the extent that the guidelines do not give effect to the intention of the parties they should not, presumably, be applied. Secondly, their Lordships appeared to recognise that there are some contexts, such as the fact situation in *HIH Casualty and General Insurance* itself, where the courts will more readily infer that the intention of the parties, or the purpose behind their contract structure, was to entitle one party to exclude or limit liability for his own negligence or the negligence of those who act on his behalf. But this is not to say that the guidelines set out by Lord Morton are to be discarded. On the contrary, they have been retained and they will continue to be applied by the courts when they appear to give effect to the intention of the parties. That this is so can be seen from the speech of Lord Bingham when he stated that

> There can be no doubting the general authority of [Lord Morton's principles], which have been applied in many cases, and the approach indicated is sound. The courts should not ordinarily infer that a contracting party has given up rights which the law confers upon him to an extent greater than the contract terms indicate he has chosen to do; and if the contract terms can take legal and practical effect without denying him the rights he would ordinarily enjoy if the other party is negligent, they will be read as not denying him those rights unless they are so expressed as to make clear that they do.[46]

This attempt to retain a role for the *Canada Steamship* guidelines is unlikely to command universal assent. The Australian courts have gone further and have departed from the *Canada Steamship* guidelines. In *Schenker & Co (Aust) Pty Ltd v Malpas Equipment and Services Pty Ltd*[47] McGarvie J concluded that

[42] [2003] UKHL 6.
[43] Para [11].
[44] *Ibid*.
[45] See Lord Bingham at para [11], Lord Hoffmann at paras [61]–[63], Lord Hobhouse at para [95] and Lord Scott at para [116].
[46] Para [11].
[47] [1990] VR 834.

> to construe commercial contracts as they would be understood by business people serves primary aims of both the law and commerce. The law serves the community best if citizens understand it and are able to resolve their dispute themselves by reference to it, without resorting to lawyers.[48]

The first sentence very much reflects the sentiments of Lord Hoffmann's approach in *Investors Compensation Scheme* and, in principle, there is much to be said for it. But it is a route that must lead to the over-ruling of the *Canada Steamship* line of authority and, so far, the House of Lords has not been prepared to take that step. Were the courts to take that step and depart from the *Canada Steamship* guidelines, this would leave the regulation of exclusion clauses largely[49] to the control of the Unfair Contract Terms Act 1977.

The potential impact of Lord Hoffmann's approach on the courts' approach to the interpretation of exclusion clauses can be further seen in *British Fermentation Products Ltd v Compair Reavell Ltd*[50] where the court adopted a more relaxed approach towards the construction of the relevant exemption clause and the principal aim of the judge was to give effect to the commercial purpose of the transaction. The defendants supplied the claimants with an air compressor for a price of just under £300,000. The claimants alleged that the compressor did not perform according to its contractually guaranteed level. They brought a claim for damages of over £1 million for the increased operating costs for the life of the machine. The defendants relied on an exclusion clause by way of defence. In essence the clauses stated that, in the event of the goods proving to be defective, the remedy of the buyer was to return the goods to the seller who would then repair or replace the defective goods and that the seller would not have any other liability in respect of defects in the goods. Judge Bowsher held that the clause was effective to exclude liability on the part of the defendants. He set out Lord Hoffmann's principles in full and concluded

> the business common-sense intention of the agreement as a whole is that the vendors undertake to supply a machine of the specification warranted, and if they fail in that undertaking the purchasers have an initial right to withdraw from the contract and reject the machine on terms that the vendors pay for them to buy from other suppliers a machine that is up to specification. If the purchasers so choose, there will be a period when the vendors will try to bring the machine up to specification, and those efforts again may be terminated by

[48] *Ibid*, at p 846.
[49] The word 'largely' is probably still necessary because in the case of attempts to exclude liability for fraud there may still be a need for a special rule which denies to a party the possibility of excluding liability for his own fraud (whether this rule is a rule of law or of construction remains a matter for debate: see *S Pearson & Son Ltd v Dublin Corporation* [1907] AC 351 and *HIH Casualty and General Insurance Ltd v Chase Manhattan Bank* [2003] UKHL 6).
[50] [1999] BLR 352.

the purchasers by rejecting on the same agreed terms. If the purchasers still do not reject when the machine fails to come up to specification, the purchasers keep the machine but on terms that they do not complain thereafter of the failure to come up to specification. The amount of the damages claimed in this action compared with the purchase price shows the good business common sense of the contract. If the project is not successful, the purchasers have two opportunities to withdraw and buy a substitute machine of the standard warranted from another supplier at the vendor's expense, but they are not to be allowed to stand on the deal and charge the vendors enormous sums for their loss continuing for the life of the machine.[51]

A further indication of the changes that are taking place can be seen in the approach taken by the courts to the construction of clauses that purport to give to one party the right to terminate the contract. The current edition of Chitty on *Contracts* states that 'where reliance is placed [when terminating the contract] on the procedure laid down in the contract, it is necessary to comply strictly with the procedure which has been laid down.'[52] But in *Ellis Tylin Ltd v Cooperative Retail Services Ltd*[53] Judge Bowsher, applying the analysis of the House of Lords in *Mannai Investment Co Ltd v Eagle Star Life Assurance Co Ltd*[54] and *Investors Compensation Scheme*, concluded that, when interpreting a clause in a contract which lays down a procedure for termination of the contract, the court will have regard to the commercial purpose that is served by the termination clause and interpret it in the light of that purpose, with the result that strict or precise compliance with the termination clause may no longer be a necessary pre-requisite to a valid termination.

C. The Role of Precedent

The starting point is that the role of precedent in cases concerned with the interpretation of contracts is limited. As Dillon L J observed in *Surrey Heath Borough Council v Lovell Construction Ltd*[55] 'a decision on a different clause in a different context is seldom of much help on a question of construction.'[56] The weight to be given to precedents has been an issue of some controversy post-*Investors Compensation Scheme*. The conflict of judicial opinion was apparent in the decision of the House of Lords in *Bank of Credit and Commerce International SA v Ali*.[57] Lord Hoffmann in his dissenting speech stated:

[51] *Ibid*, at p 358.
[52] 28th edn, (1999) para 23–047.
[53] [1999] BLR 205.
[54] [1997] AC 749.
[55] (1990) 48 Build LR 113.
[56] *Ibid*, at p 118.
[57] [2001] UKHL 8; [2002] 1 AC 251.

> If interpretation is the quest to discover what a reasonable man would have understood specific parties to have meant by the use of specific language in a specific situation at a specific time and place, how can that be affected by authority?[58]

Lord Bingham, speaking for the majority, accepted that 'the authorities must be read in the context of their peculiar facts'[59] but he nevertheless had regard to the authorities for the purpose of establishing the principle that the courts are reluctant to infer that a party intended to give up something which neither he, nor the other party, knew or could know he had. It would therefore be going too far to conclude that the authorities are irrelevant when seeking to interpret a contract. But they should not be given undue weight.[60] In *Midland Bank plc v Cox McQueen*[61] Mummery L J stated:

> Detailed comparisons of one document with another and of one precedent with another do not usually help the court to reach a decision on construction. Indeed, that exercise occupies a disproportionate amount of valuable time which would be better spent on the arguments that really count: those which focus on the precise terms of the relevant documents and the illuminating environment of the transaction.[62]

Matters are rather different in the case where a word or phrase has built up an established meaning in the case-law. In such a case the precedents assume greater significance and the courts are likely to give the words their established meaning. Thus in *British Sugar plc v NEI Power Projects Ltd*[63] Waller L J stated, when considering the meaning of the words 'consequential loss', that

> once a phrase has been authoritatively construed by a court in a very similar context to that which exists in the case in point, it seems to me that a reasonable businessman must more naturally be taken to be having the intention that the phrase should bear the same meaning as construed in the case in point. It would again take very clear words to allow a court to construe the phrase differently.[64]

[58] *Ibid*, at para [51].
[59] *Ibid*, at para [17].
[60] *Daejan Properties Ltd v Bloom*, 30 June 2000, Court of Appeal; *Midland Bank plc v Cox McQueen* [1999] 1 FLR 1002; *Gibbs v Ebbetts*, 26 October 1997, Court of Appeal. See *Pamment v Sutton*, 13 November 1998, Chancery Division where Deputy Judge Peter Whiteman QC derived some guidance from the authorities.
[61] [1999] 1 FLR 1002.
[62] *Ibid*, at p 1012.
[63] (1997) 87 BLR 42.
[64] *Ibid*, at p 50. That said the meaning of the phrase 'indirect or consequential loss' has been thrown into doubt by the observation of Lord Hoffmann in *Caledonia North Sea Ltd v British Telecommunications plc* [2002] UKHL 4, [2002] 1 Lloyd's Rep 553 where he stated (at para [100]) that he wished to 'reserve the question' whether the Court of Appeal had adopted the correct construction of the phrase 'indirect or consequential loss.'

But in other cases there is no substitute for a careful examination of the particular words and phrases that have been used by the parties to the particular contract.

D. Has There Been a Revolution?

A consideration of the impact of Lord Hoffmann's re-statement on existing precedents leads on to a related question of the extent of the change that has taken place in English law. How much of a change have Lord Hoffmann's principles introduced? As has been noted, Lord Hoffmann did suggest that there has been a 'fundamental change' in the law but he did not attribute that change to his own judgment. Rather, he attributed it to the judgments of Lord Wilberforce in *Prenn v Simmonds*[65] and *Reardon Smith Line Ltd v Hansen-Tangen*.[66] The courts have on two occasions rejected the submission that Lord Hoffmann's re-statement constituted a revolution in the law.[67] It is probably true to say that the law is slowly evolving. It is evolving both in relation to the range of materials which the court is willing to admit into evidence and in relation to the extent to which it is prepared to depart from the literal meaning of the words used by the parties and adopt a construction which leads to a commercially sensible construction.

E. The Admissibility of Pre-contractual Negotiations

The evolution that is taking place is, in certain respects at least, a slow one. This is particularly so in relation to the admissibility of evidence of pre-contractual negotiations. Lord Hoffmann, in his third principle, stated that evidence of pre-contractual negotiations is generally inadmissible, although he did note that the exceptions to the rule are 'in some respects unclear'. The issue of the admissibility of pre-contractual negotiations has arisen in 24 cases post-*Investors Compensation Scheme* cases and the courts have, in general, been content to affirm the general exclusionary rule.[68]

[65] [1971] 1 WLR 1381.
[66] [1976] 1 WLR 989.
[67] *Cameron v M & W Mack (ESOP) Trustee Ltd*, unreported, 8 November 2001; *New Hampshire Insurance Co v Philips Electronics North America Corporation (No 2)* [1999] LRLR 66, 70.
[68] See, for example, *John v Price Waterhouse*, 24 June 2002, Court of Appeal; *Aqua Design & Play International Ltd (in liquidation) v Kier Regional Ltd*, [2003] BLR 111, *Jones v Forest Fencing Ltd*, 20 November 2001, Court of Appeal; *Bank of Credit and Commerce International SA v Ali* [2001] UKHL 8; [2002] 1 AC 251, para [31]; *P & O Overseas Holdings Ltd v Rhys Braintree Ltd*, 5 July 2001, Chancery Division; *Champion v Workman*, 20 July 2001, Chancery Division; *John v Price Waterhouse*, 11 April 2001, Chancery Division; *Sloggett and Perry Ltd v Stroud*, 25 May 2000, Court of Appeal; *Nella v Nella*, 23 July 1999, Chancery Division; *Inland Revenue Commissioners v Botnar*, Court of Appeal, 23 June 1999.

The most recent consideration of the issue was provided by the Privy Council, on appeal from the Court of Appeal of New Zealand, in *Canterbury Golf International Ltd v Yoshimoto*.[69] The judgment of the Privy Council was delivered by Lord Hoffmann. In the Court of Appeal of New Zealand Thomas J had expressed the view that his construction of the contract was supported by two provisions in earlier drafts of the contract and he further stated that the rule that excluded evidence of pre-contractual negotiations should be departed from or relaxed. In the Privy Council Lord Hoffmann stated that the case did not provide a 'suitable occasion' for a re-examination of the law because on the facts the evidence of the pre-contractual negotiations was 'unhelpful.' He stated that

> Their Lordships do not think that it is helpful to try to construe the earlier version of clause 6.3 because it was dropped and the present clause 6.3 substituted. It seems to them pointless to try to speculate upon why the change was made. No doubt each party had their reasons for proposing it on the one hand and accepting it on the other. All a court can do is decide what the final contract means.[70]

Why does the law exclude from consideration evidence of pre-contractual negotiations? A number of possible reasons can be given. The first is that the test to be applied when seeking to ascertain the existence of a contract and its terms is an objective one, not subjective. The court should not therefore concern itself with subjective statements by either party as to the meaning to be ascribed to the particular term. The law does not wish to encourage parties to make self-serving statements in the course of negotiations and then produce them in evidence when a dispute breaks out in relation to the meaning of that particular term. Secondly, it is objected that evidence of pre-contractual negotiations is not helpful on the basis that the parties' position is constantly changing prior to the conclusion of the contract and the court should not be asked to speculate on the possible reasons for the changes that have been made on the way to concluding the contract. The task of the court is to ascertain the meaning of the contract, not the process that led up to the formation of the contract. But these arguments are not all one way. In the first place it can be argued that these are objections that relate to the weight of the evidence, not its admissibility. Second, in so far as the courts now state that they should seek to adopt a construction which advances the commercial purpose which the parties had in mind, is it not the case that the pre-contractual negotiations may provide very good evidence as to what was the commercial purpose of the parties?

The exclusion of pre-contractual negotiations is not, however, absolute. Where there is an ambiguity in the final written document, evidence of

[69] [2002] UKPC 40.
[70] Para [28].

pre-contractual negotiations may be admissible to show that the parties had attached a particular meaning to that phrase.[71] Similarly, as Lord Hoffmann acknowledged, such evidence is admissible in an action for rectification. Further, it would appear that the position is that evidence of background and of the object and genesis of the transaction is admissible but that evidence of the parties' previous negotiating positions is not.[72] This distinction is not an easy one to draw. As Judge MacDuff QC stated in *Sykes v Pannell Kerr Forster*

> there is … a very fine borderline between the admissible and the inadmissible. One may look at the facts and circumstances surrounding the negotiation and that about which the parties were negotiating, but not at what the parties said during negotiations nor at previous drafts. The rationale, at least in part, appears to be a wish on the part of the courts to encourage open negotiations and to discourage the creation of self-serving documents or utterances.[73]

Judges have on a number of occasions noted the difficulty in distinguishing in this context between evidence which is admissible and evidence which is not.[74] Now that the courts give more weight to the adoption of an approach which seeks to give a commercially sensible construction to the clause in dispute, the justification for excluding pre-contractual negotiations from evidence appears suspect. They may in fact provide very good evidence of the issue that was at stake between the parties and, to that extent, should be admissible in evidence. The courts should continue to exclude from evidence 'self-serving utterances' but the reason for this exclusion is that it is for the court, not the parties, to decide what the contract means. Lord Hoffmann's first principle makes the point that the test applied by the court is an objective, not a subjective, one. Evidence of the subjective state of mind of the parties is therefore irrelevant and this is so whether the evidence relates to his or her state of mind during the negotiations or at a later stage in the process. This being the case, evidence of pre-contractual negotiations should be admissible in evidence unless that evidence relates to the subjective state of mind of the negotiating parties. This is not to say that a great deal of weight

[71] *The Karen Oltman* [1976] 2 Lloyd's Rep 708.

[72] *Gehe AG v NBTY Inc*, unreported, QBD, 30 July 1999, *cf The Management Corporation Strata Title Plan No 193 v Liang Huat Aluminium Ltd* [2001] BLR 351.

[73] Unreported, QBD, 30 March 2001. See also *Bank of Scotland v Dunedin Property Investment Co Ltd* 1998 SC 657 where Lord President Rodger stated that the rationale behind the exclusionary rule 'shows … that it has no application when the evidence of the parties' discussions is being considered, not in order to provide a gloss on the terms of the contract, but rather to establish the parties' knowledge of the circumstances with reference to which they used the words in the contract'.

[74] See, for example, *Gould v BG Transco Ltd*, unreported, Chancery Division, 10 August 2001; *The Tychy (No 2)* [2001] EWCA Civ 1198; [2001] 2 Lloyd's Rep 403; *Biggin Hill Airport Ltd v Bromley London Borough Council*, *The Times*, 9 January 2000; *Gehe AG v NBTY Inc*, unreported, QBD, 30 July 1999 and *Demolition Services Ltd v Castle Vale Housing Action Trust*, (1999) 79 Con LR 55.

should necessarily be given to evidence of pre-contractual negotiations. The weight will very much depend upon the facts of the case. But in principle evidence of pre-contractual negotiations should be admitted, even if, in many cases, it ultimately proves to be 'unhelpful'.

F. The Admissibility of Evidence of Conduct Subsequent to the Making of the Contract

Lord Hoffmann makes no express reference in his speech in *Investors Compensation Scheme* to the admissibility of evidence of conduct subsequent to the making of the contract. Pre-*Investors Compensation Scheme* authority establishes that such evidence is inadmissible,[75] although it may be relevant to a plea of estoppel, including estoppel by convention,[76] and to the question whether a contract has been subsequently varied.[77] The reason given for the exclusion of this evidence is that, were it admissible, the contract could mean one thing on the day on which it was signed but mean something completely different one month after it was signed by virtue of the conduct of the parties after the making of the contract. The exclusion of this evidence has been affirmed in a number of cases post-*Investors Compensation Scheme*.[78]

G. When can the Courts Depart from the Natural and Ordinary Meaning of the Words Used?

One of the issues which has arisen in the cases post-*Investors Compensation Scheme* is the point in time at which it is legitimate for a court to have regard to the commercial purpose of the transaction or to consider whether or not the parties have used words in an unnatural sense. Is it only when the words used by the parties are ambiguous or is it in all cases? Lord Hoffmann's re-statement appears to suggest that it is in all cases. Indeed, in *Westminster City Council v National Asylum Support Service*[79] Lord Steyn stated that Lord Hoffmann had made 'crystal clear that an ambiguity need

[75] *Schuler AG v Wickman Machine Tool Sales* [1974] AC 235.
[76] *James Miller & Partners Ltd v Whitworth Street Estates (Manchester) Ltd* [1970] AC 583; *Mannai Investment Co Ltd v Eagle Star Life Assurance Co Ltd* [1997] AC 749, 768 (Lord Steyn) and 779 (Lord Hoffmann); *The Tychy (No 2)* [2001] EWCA Civ 1198, [2001] 2 Lloyd's Rep 403.
[77] *Philip Collins Ltd v Davis* [2000] 3 All ER 808, 822.
[78] *Carillion Construction Ltd v Farebrother and Partners (a firm)*, 21 February 2002, Queen's Bench Division; *The Tychy (No 2)* [2001] EWCA Civ 1198, [2001] 2 Lloyd's Rep 403; *Bank of Credit and Commerce International SA v Bugshan*, 16 February 2001, Court of Appeal; *Daejan Properties Ltd v Bloom*, 30 June 2000, Court of Appeal; *Philip Collins Ltd v Davis* [2000] 3 All ER 808, 822; *Demolition Services Ltd v Castle Vale Housing Action Trust* (1999) 79 Con LR 55; and *Robertson v Hill Samuel Investment Services Ltd*, 14 January 1999, Employment Appeal Tribunal.
[79] [2002] UKHL 38.

not be established before the surrounding circumstances may be taken into account.'[80] It is the breadth of this statement, and the uncertainty that it may engender, that has attracted the criticism of judges such as Saville L J in *Bank of Sharjah v Dellborg*.[81]

In this context it may be important to recall that *Investors Compensation Scheme* was a case in which the parties were held to have used words in an unnatural sense and so it was necessary for the court to have regard to the surrounding circumstances in order to ascertain the intention of the parties in relation to the words they had used. *Investors Compensation Scheme* was not a case in which the words had been used by the parties in a clear and natural sense and so it was not necessary for Lord Hoffmann to address his mind to the case in which the parties use words in their natural, ordinary meaning but one party wishes to adduce evidence of surrounding circumstances in order to demonstrate that the parties have used the words in a secondary or unnatural sense. In cases where it is not apparent that the parties have used words in an unnatural sense, and the document has been drawn up with professional advice, there is evidence of judicial reluctance to have regard to the surrounding circumstances.[82] Thus in *Breadner v Granville-Grossman*[83] Park J stated that 'it remains the case that the starting point, and usually the finishing point as well, is to identify the natural and ordinary meaning of the words which the draftsman has used.'[84] In some cases the judges have given a meaning to the words used in the documents without regard to the surrounding circumstances and then used these circumstances as a check on the conclusion they have already reached.[85]

H. A Purposive Approach?

What is the relationship between Lord Hoffmann's principles and the more general trend, noted in particular in a number of Lord Steyn's speeches,[86] towards the adoption of a purposive approach to the interpretation of documents? Lord Hoffmann's re-statement makes no reference to the adoption

[80] Para 5.
[81] Unreported, Court of Appeal, 9 July 1997, on which see further n 15 above and associated text.
[82] See, for example, *Pandora Investments SA v Momentum Pty Ltd*, Court of Appeal, 17 December 1999.
[83] [2001] Ch 523. See also *Henry Boot Construction Ltd v Alstom Combined Cycles Ltd* [2000] BLR 247 and *Pandora Investments SA v Momentum Pty Ltd*, Court of Appeal, 17 December 1999.
[84] [2001] Ch 523, para [36].
[85] See, for example, *Columbia Tristar Home Video (International) Incorporated v Polygram Film International BV*, 8 February 2000, Court of Appeal; *Demolition Services Ltd v Castle Vale Housing Action Trust* (1999) 79 Con LR 55.
[86] See, in particular, *Deutsche Genossenschaftsbank v Burnhope* [1995] 1 WLR 1580, 1589; *Lord Napier and Ettrick v R F Kershaw Ltd* [1999] 1 WLR 756, 763; *Mannai Investment Co Ltd v Eagle Star Life Assurance Co Ltd* [1997] AC 749, 770 and *Total Gas Marketing Ltd v Arco British Ltd* [1998] 2 Lloyd's Rep 209, 221.

of a purposive approach to the interpretation of a document, although his fifth principle does contain an important reference to a construction that promotes 'business common sense', a theme that is prominent in the speeches of Lord Steyn. The focus of Lord Hoffmann's analysis is upon the words used, their meaning and the material to which a court can have regard when seeking to ascertain the meaning of the words used. Nevertheless, his re-statement has been regarded as authority in subsequent cases for the adoption of a 'purposive construction of commercial agreements.'[87] It is not clear that this is correct. Lord Hoffmann's primary emphasis is on the meaning of words, not the purpose that the parties had in mind when entering into the transaction. Their purpose is only relevant in so far as it can shed light on the meaning of the words used by the parties. The courts cannot substitute different words than those used by the parties simply because the substituted words 'would satisfy the commercial objective equally well or better.'[88]

I. The Matrix of Fact

The use of the language of 'matrix of fact' has been attacked, most notably by Sir Christopher Staughton.[89] He has criticised the use of language such as 'the matrix of fact' on the ground that 'counsel have wildly different ideas as to what a matrix is and what it includes.' Particularly difficult was the use by Lord Hoffmann of the words 'absolutely anything' in his second principle. These words appeared likely to encourage lawyers to seek to adduce evidence that was previously inadmissible by introducing it under the guise of the 'matrix of fact'. Lord Hoffmann has since 'qualified' his second principle. In *Bank of Credit and Commerce International v Ali*[90] he stated that:

> I did not think it necessary to emphasise that I meant anything which a reasonable man would have regarded as *relevant*. I was merely saying that there is no conceptual limit to what can be regarded as background. It is not, for example, confined to the factual background but can include the state of the law (as in cases in which one takes into account that the parties are unlikely to have intended to agree to something unlawful or legally ineffective) or proved common assumptions which were in fact quite mistaken.[91]

[87] See, for example, *Crest Homes (South West) Ltd v Gloucestershire County Council*, unreported, Court of Appeal, 22 June 1999 per Nourse LJ and *Ellis Tylin Ltd v Co-operative Retail Services Ltd* [1999] BLR 205.

[88] *Canterbury Golf International Ltd v Yoshimoto* [2002] UKPC 40, para [18].

[89] 'How do the Courts Interpret Commercial Contracts?' [1999] *Cambridge Law Journal* 303 and for similar views expressed in a judicial capacity see *Scottish Power plc v Britoil (Exploration) Ltd, The Times*, 2 December 1997.

[90] [2001] UKHL 8, [2002] 1 AC 251.

[91] *Ibid*, at para [39].

The cases post-*Investors Compensation Scheme* make reference to the 'matrix of fact' but do not suggest that the judges have experienced particular difficulty in ascertaining what does and what does not fall within the scope of the matrix of fact.[92] The uncertainty seems to relate to the weight to be given to the matrix of fact when weighed against the natural and ordinary meaning of the words rather than the definition of the matrix of fact itself.

J. 'Commercial Construction' and 'Re-writing' the Contract

One problem that has arisen in the case-law relates to the situation in which one party alleges that the contract contains a drafting mistake. In such a case can the court depart from the ordinary meaning of the words used in order to give effect to what is alleged to be the purpose behind the clause in question? Cases can be found in which the courts have disregarded some of the words in a contract in order to give effect to the purpose which the court believes the parties had in mind when entering into the contract. A good example in this connection is provided by the decision of the Court of Appeal in *Sinochem International Oil (London) Co Ltd v Mobil Sales and Supply Corporation*[93] where the court did make some allowance for the 'vicissitudes of drafting'[94] in order to give effect to the 'general aim' of the contract. Thorpe L J stated that cases such as *Investors Compensation Scheme* 'allowed a sufficient latitude in construction to justify the rejection of what [was] claimed to be the literal meaning of the disputed clause.'[95] Similarly it has been stated that the court may 'require what appear to be errors or inadequacies in the choice of language to yield to [the intention of the parties] and be understood as saying what (in the light of that purpose) that language must reasonably be understood to have been intended to mean.'[96] The difficulty with this approach is that it may cross the line between a 'commercial construction' of the contract and 're-writing' the contract between the parties.[97]

The difficulty is that the line between 'commercial construction' (which is permissible) and 're-writing' the contract (which is not) is a difficult one to draw.[98] The difficulty can be seen in *Investors Compensation Scheme*

[92] See, for example, *Pamment v Sutton, The Times*, 15 December 1998; *C Itoh & Co Ltd v Companhia De Navegacao Lloyd Brasileiro and Steamship Mutual Underwriting Association (Bermuda) Ltd (The 'Rio Assu') (No 2)* [1999] 1 Lloyd's Rep 115.
[93] [2000] 1 Lloyd's Rep 339
[94] *Ibid*, at p 345.
[95] *Ibid*, at p 346.
[96] *Don King Productions Ltd v Warren* [1998] 2 All ER 608, 624, cited with approval by Morritt L J in the Court of Appeal, [1999] 2 All ER 218, 229–230.
[97] *Sinochem International Oil (London) Co Ltd v Mobil Sales and Supply Corporation* [2000] 1 Lloyd's Rep 339, 340 per Mance LJ.
[98] See, for example, Gee 'Construing Contracts with Mistakes in Them' [2001] *Lloyd's Maritime and Commercial Law Quarterly* 214, criticising the approach of the Court of Appeal in *Hombourg Houtimport BV v Agrosin Private Ltd (The Starsin)* [2001] 1 Lloyd's Rep 437.

itself where Lord Lloyd dissented on the basis that the majority had crossed the line between a purposive interpretation and creative interpretation. The courts have traditionally been unwilling to adopt a construction that leads to a very unreasonable result.[99] This approach to interpretation reflects the ordinary perception that contracting parties are unlikely to have agreed to something absurd. But this rule of construction has its limits, difficult though it may be to find them. As Lord Mustill stated in *Charter Reinsurance Co Ltd v Fagan*,[100] there

> comes a point at which the court should remind itself that the task is to discover what the parties meant from what they have said, and that to force upon the words a meaning which they cannot fairly bear is to substitute for the bargain actually made one which the court believes could better have been made. This is an illegitimate role for a court.[101]

IV. CONCLUSION

Lord Hoffmann's re-statement has clearly had a considerable impact on English law, at least if impact is to be judged by the number of citations. More difficult to quantify is the impact of his re-statement on the content of the substantive rules of law. The courts today are probably readier than their predecessors to have regard to the surrounding circumstances of the contract when seeking to ascertain the meaning of the contract or of a particular term. But the courts remain generally unwilling to admit evidence of pre-contractual negotiations or of conduct subsequent to the conclusion of the contract. The impression that is given is that the law is evolving but there has, as yet, been no revolution.

How will the law develop in the future? One possible development is that the law might adopt an approach similar to that taken in Articles 5.101 and 5.102 of the Principles of European Contract Law (which were not cited in any of the cases considered in this study). Article 5.101 provides:

> (1) A contract is to be interpreted according to the common intention of the parties even if this differs from the literal meaning of the words.
> (2) If it is established that one party intended the contract to have a particular meaning, and at the time of the conclusion of the contract the other party could not have been unaware of the first party's intention, the contract is to be interpreted in the way intended by the first party.
> (3) If an intention cannot be established according to (1) or (2), the contract is to be interpreted according to the meaning that reasonable persons of the same kind as the parties would give to it in the same circumstances.

[99] See *Schuler AG v L Wickman Machine Tool Sales* [1974] AC 235.
[100] [1997] AC 313.
[101] *Ibid*, at p 388.

This Article does not appear to differ substantially from Lord Hoffmann's re-statement. Article 5.101(1) is likely to generate a degree of unease among English lawyers but no more than that aroused by Lord Hoffmann's fourth and fifth principles. Article 5.102 is slightly more contentious. It states:

> In interpreting a contract, regard shall be had, in particular, to:
> (a) the circumstances in which it was concluded, including the preliminary negotiations;
> (b) the conduct of the parties, even subsequent to the conclusion of the contract;
> (c) the nature and purpose of the contract;
> (d) the interpretation which has already been given to similar clauses by the parties and the practices they have established between themselves;
> (e) the meaning commonly given to terms and expressions in the branch of activity concerned and the interpretation similar clauses may already have received;
> (f) usages; and
> (g) good faith and fair dealing.

Paragraphs (a), (b) and (g) seem clearly to go beyond the current limits of English law. Those who maintain that Lord Hoffmann has not gone far enough argue that English law should embrace propositions (a) and (b) so that courts will in future be free to assess for themselves the probative value of such evidence (which may not be great). On the other hand, care must be taken not to lengthen trials by enabling the parties to swamp the court with evidence of dubious value. Paragraph (g) is also of interest. English contract law currently does not impose on contracting parties a duty of good faith and fair dealing. However good faith and fair dealing play a vital role in civilian systems. A huge gulf thus appears to exist between English law and Continental systems. But the difference may be more one of technique than outcome. As Lord Hoffmann observed in *O'Neill v Phillips*,[102] the result which an English court might achieve by adopting a less literal approach to interpretation might well by reached in a Continental court by the use of a general requirement of good faith. So, at the end of the day, they may turn out to be no more than 'different ways of doing the same thing.' We may yet move in the direction pointed by Article 5–102 of the Principles.

[102] [1999] 1 WLR 1092, 1101.

7

The Uses of Ambiguity in Commercial Contracts: On Facilitating Re-Bargaining

WILLIAM T ALLEN AND GALYA LEVY

I. INTRODUCTION

THIS CHAPTER ASKS why we observe in various common law jurisdictions the judicial invocation of an evolutionary concept: the implied duty of good faith in contracts between parties who under traditional principles would not be considered fiduciaries for each other. This is puzzling if, with some American commentators, we tend to think (1) that an efficient contract law is one that enforces the ascertainable intent of the contractual parties and minimises imposition through the contract of rights or duties other than those agreed upon; and (2) that the common law of contracts tends to evolve in a way that makes sense (is efficient).

We concur in the conventional view that clarity in legal rights and duties is one of the principal goals of contract law. Clear rules allow for the creation of clear obligations and permit parties to best understand and price the performances about which they contract. Accurate pricing of contract performances in turn aids the achievement of an efficient allocation of goods, services and the capital that supports the production of goods and services. Thus ambiguity in rights and duties can be costly. But we wish in this essay to note a small but we hope interesting caveat.

In this chapter we wish to argue that in some circumstances clarity in the fixing of default legal rights may not improve contract efficiency. Ambiguity has its uses. It is not always simply a flaw or a regrettable feature of a legal system, but can sometimes be a productive state. We undertake to demonstrate this through a discussion of the effects of the ambiguity that inevitably arises from the doctrine of an implied covenant of good faith and fair dealing.

A. The Implied Duty to Act Fairly Towards a Counter-Party

Over the last half century the jurisprudence in several common-law countries has embraced in one way or another, and to one degree or another, the general notion that at least in contracts in which the parties have an on-going interaction, each party owes to the other an implied obligation of good faith.[1] American law of contracts has perhaps been the most aggressive in this view. The statutory law of sales of goods includes an express mandate providing that all sales of goods contain an unwaivable obligation of 'good faith and fair dealing.'[2] Moreover the Restatement of the Law of Contracts, Second, which purports to restate the existing general law of contracts in the US, asserts that such a duty is implied in law in all contracts.[3] Undoubtedly common law jurisdictions across the world have over the last century moved, if unevenly and sporadically, from the liberal world of *caveat emptor* to a world in which concerns for bargaining power and gamesmanship sometimes motivate courts to be more active in fashioning 'just' judicial results. Noting this evolution early on, Professor Grant Gilmore, famously, if rather prematurely as it turned out, foresaw the 'death of contract.'[4] While Professor Gilmore's foresight in fact exaggerated the extent to which US courts would move from traditional concepts of contract to the broad tort-like protection of the reliance interest, he was correct in observing a judicial tendency to be more active in shaping 'fair' remedies in contract cases.

a. Loyalty in Commercial Relations

The notion that parties in a relationship might owe to each other a general duty to act with some degree of loyalty to their common undertaking derives from the holy place of English law: the sanctum in which the Chancellor shaped the duty of fiduciaries. The duty of loyalty there constructed in order to govern the duty that a trustee owes to the trust and to his *cestui que trust* is the strictest form of propriety the common law system observes. No act implicating a personal interest of the trustee is permissible, unless expressly sanctioned in the document creating the trust.[5] Even in the less strictly monitored area of fiduciary duties of corporate directors, the exercise of judgment by members of a corporation's board of directors may be closely monitored by a court of equity to ensure that directors act in

[1] Eg, *Shell Oil v Marinello* 294 A 2d 253 (1972); *Dayan v McDonald's Corp* 466 NE 2d 958 (Ill App. 1984); *Scheck v Burger King Corp* 756 F Supp 543 (SD Fla 1991); *Gallagher v Pioneer Concrete (NSW) Pty Ltd* [1993] ATPR 41–216 (Australia); *Re Empress Towers v Bank of Nova Scotia* (1991) 71 DLR 400 (Canada).
[2] See *Uniform Commercial Code*, § 2–203.
[3] American Law Institute, *Restatement of the Law of Contracts*, *Second* § 295 (1981).
[4] Grant Gilmore, *The Death of Contract* 2nd edn (Ronald K L Collins (ed), (1995).
[5] American Law Institute, *Restatement of the Law of Trusts*, *Second* (1959).

a good faith attempt to advance the corporate purpose. They may never seek to advance a personal interest at a cost to the corporation.[6] The fiduciary duty of loyalty serves an important economic function in those fields of law in which it plays a part. It encourages relationships (and investments), both personal and business, in which one party vests trust in another. As such, the nature and extent of the fiduciary obligation is quite important as a default obligation in the law of corporations, facilitating the ability of firms to raise large volumes of capital from large numbers of investors through public equity markets. The heart of this storied obligation (no source of obligation leads the work-a-day judge to the plane of grandiloquence more readily) is the requirement that a fiduciary exercise *good faith* judgment in an effort to advance the purposes of the institution (the trust, the estate, the corporation) he serves. Thus it is a legal obligation which, at the end of the day, is of striking breadth and generality. Inevitably, opinions concerning fiduciary duties involve a painstaking evaluation of myriad facts and circumstances, followed by a *ex post hoc* judgment concerning the actor's conformity with some radiant generality.

But this exacting standard of selflessness or devotion has fittingly been restricted to instances in which the defendant holds property or power over property that another person owns 'equitably'. We may well ask, then, in what way do we imagine that the importation into the hubbub of commercial law of a diluted version of the duty of loyalty—for that may be a way to describe the implied duty of good faith and fair dealing—will increase human welfare? Is this innovation likely over time to increase social wealth by facilitating contracting behaviour? Or is this innovation simply a fuzzy-headed (costly) reflection of an ethic of post hoc re-distribution based upon moral intuitions that, while perhaps socially attractive to some, would not be sustained were we to be able (as we are not) to subject it to a realistic cost benefit analysis?

How to explain this evolution in which fuzzy concepts come to play a role in contract construction? For those who find functional explanations of legal phenomena satisfying, a merely descriptive account of an evolutionary morality affecting contract law seems inadequate. Such persons prefer to understand the economic and technological forces that shape such trends. High in the hierarchy of functionalist tools for understanding are the set of assumptions that underlie neo-classical economics (eg, self-seeking motivation of individual action, human rationality, and the inherent subjectivity of value). To these assumptions can be added a further set that attempt to incorporate more of the features of actual human life: information asymmetry, human guile and worse, the boundedness of human rationality, the pervasiveness of 'agency costs' in complex contracting and the costliness of contracting. Any economic explanation of how legal institutions work,

[6] *SIPCA Holdings SA v Optical Coating Laboratory, Inc.*, 22 *Del. J Corp. L* 1282.

or how they ought to be fashioned in order to work well, that incorporate some or all of these additional assumptions should, we imagine, be more helpful in the practical work that lawyers and judges do.

Scholars in the US at least, who seek to use economic rationality as an analytic technique in understanding legal institutions—either the neo-classical version or the more realistic version—take a stance regarding change in legal institutions that looks vaguely Darwinian. It sees the common law system with its relatively strong commitment to human liberty and decentralised (judicial) law creation as permitting great experimentation. To such observers, decentralised common law systems permit gradual system change towards a more satisfactory (efficient) set of rules. Inefficient rules may emerge of course, but the costly aspects of the rule become apparent over time and sytemtic adaptation occurs. It is thought that in time an efficient rule will be stable (while its environment is stable) and the system will move towards such rules. To a functionalist, the existence of an established legal pattern is an implicit invitation to imagine in what way it serves an economising purpose. This essay is in that spirit.

We want to offer an efficiency-based account of the evolution of the 'good faith' term implied in modern US contracts. Thus, we offer a functional account of why and in which circumstances the implication of this duty would be useful to the parties' joint interest at the time of contracting (*'ex ante'*). To produce the 'bottom line' immediately we suggest that the implied doctrine of good faith is useful in long-term contracts precisely because it is ambiguous. This ambiguity, we assert, can facilitate *ex post* re-bargaining of a contract after an unknown future state is revealed to the contracting parties and it can help to produce results on re-contracting that are distributionally more equal than otherwise would occur. We suggest that this indeed is the result that parties would tend to seek behind the veil of ignorance that exists at the time of contracting. Thus we assert that the ambiguity that arises from the implied obligation is an aid to contracting parties and is efficient in such cases. We do not claim that courts have restricted use of the doctrine to these circumstances, but this interpretation does, we think, suggest one way in which this doctrine (and ambiguity in contract law) can be functionally useful.

We begin by asking the reader to consider a hypothetical case of a complex contract.

II. A HYPOTHETICAL CASE

Bigco, Inc. is a manufacturer of filtration systems for industrial uses at a Texas plant. It uses a semi-refined material (stuff) in its manufacturing process. Stuff contributes about 15 per cent of the marginal manufacturing costs of Bigco's products. Over the last five years Bigco's Texas requirements for stuff fluctuated between 50,000 and 75,000 lbs each month (average 60,000).

> It has been paying an average price of \$3.40 a pound to its supplier, Ajax Corp., but Ajax has been unreliable and the relationship was terminated in 2001.
>
> That same year Bigco negotiated with Supplier Inc. Supplier proposed a different approach. It proposed to build a small refining unit right on Bigco's property at a cost of \$1.6 million which would permit it to profitably supply all of Bigco's stuff needs for average prices between \$2.40 and \$2.60 per pound. Supplier calculates that the plant would have at least an eight-year life. If it charges off the capital cost plus interest over the period the charge would be \$650,000 per year. Supplier calculates that on the expected (average) requirement of 720,000 pounds per year, it will have a 20 per cent profit margin, net of its costs of capital. Thus it foresees expected profits per year of \$360,000 (720,000 pounds × \$2.50 per pound × .20 = \$360,000).
>
> A contract is signed on 2 January 2001. It provides that Supplier will be the exclusive supplier of stuff for the Bigco plant in Texas for a period of not less than eight years on terms consistent with its proposal, with a provision for price adjustments in light of any changes in a bench market indicator of stuff prices, with a guaranteed price of at least \$2.00 per pound and a top price of no more than \$4.40 a pound, over the term of the contract. Supplier immediately commences to construct the refining unit and shortly thereafter deliveries begin.
>
> In May 2002, Gemax, a competitor of Supplier, introduced a new, advanced technology for refining stuff. The new technology improved the duration of the product's lifetime, while shortening the production time by eliminating two of the expensive stages that were used in the old refining process. They now sell stuff for \$1.60 a pound. Understanding it is bound by a contract for 'no less than eight years', Bigco asks Supplier to reduce the price for stuff to \$1.65 a pound. Supplier, who carefully calculated its expenses in accordance with the estimated requirements and prices, recognizes that it cannot afford the price reduction, and therefore refuses. Bigco is frustrated. An audit of the Texas plant operations discloses that given all of the recent changes it can expect to operate the Texas plant at no more than a \$50,000 annual profit.
>
> Bigco then announces on 1 January 2003 that it will relocate the operations of its Texas plant to a Mississippi plant which will be refitted for that purpose. Employees will be offered employment at different Bigco operations. Supplier is enraged by this decision. Its investment will have a value of \$100,000 as scrap if not used for its intended purpose at the Bigco Texas site.

Has Bigco breached any obligation that it owed to Supplier? One answer, a plausible one, is no, it has not. If Supplier needed assurance that the contract would continue for eight years in order to assure recovery of all of its investment it should have negotiated for a covenant that the plant would operate at present levels for the eight year term of the contract. This answer simply leaves the parties where they are. It is almost certainly the result that would obtain if there were no implied obligation of good faith and fair dealing. And for some of us I suppose this result does not leave us feeling as if some egregious wrong has been left unrighted.

But we suppose that what is wrong with this outcome is not that it offends our intuitions about fairness, but that it renders a better outcome less likely to be achieved. Before elaborating on why this is so and why an implied obligation and good faith would produce a better result, we wish to set the general background for our observation by discussing the efficiency standard in contract law.

A. The Efficiency Standard in Contract Law

This is old ground, more thoroughly plowed and better plowed by others,[7] but a necessary starting place for us too. To begin, we assume that in creating a law of contracts, the legal system seeks (1) to facilitate contracting activity through a variety of default rules which allow parties to economize on negotiating costs; (2) to provide reliable (disinterested and expert), expeditious and inexpensive institutions (including rules and principles) for assuring performance of contract promises and for compromising or determining and compensating claims of breach; (3) to encourage performance of promises upon which reasonable reliance has occurred, but only when such performance makes good sense (that is, is efficient in a Kaldor-Hicks sense[8]); and (4) to achieve socially defined fair results or distributions when the law's coercive power is deployed on behalf of a complaining party.

Thus among our basic assumption is the assumption that default rules of contract law are (over time) established in such a way as to *reduce costs of contracting (and thus to encourage contracting)* and to facilitate *appropriate* performance. Costs of contracting vary with the complexity of the performances contemplated and the period over which performances are foreseen. Spot market purchases of a fungible commodity for cash are probably the paradigm of a low cost contract negotiation. A requirements contract for supply of a complex production input could represent a contract of substantial complexity and cost. A franchising agreement or joint venture of an indefinite term would occupy a place near the high end of contracting costs. A part of the costs of contract negotiation is the cost of attempting to foresee future states of the world and to create explicit contract terms to deal with them. This effort requires collecting and analysing costly information and designing rights to assign and mitigate risks. As the time horizon of a

[7] Eg, Kornhauser, 'An Introduction To The Economic Analysis of Contract Remedies' (1986) 57 *U Colo L Rev* 683.

[8] Kaldor (1939) and Hicks (1939) developed the Compensation Criterion, based on the example of a reform or a switch from policy x to policy y: 'If those who gain from the switch could *in theory* compensate those who have been harmed, and remain better off, then the move is desirable, and y is better for society than x … the compensation is *not necessarily paid*; it is a theoretical possibility, not a fact.' Allan M Feldman, 'Kaldor-Hicks Compensation' in Peter Newman, (ed), 2 *The New Palgrave Dictionary of Economics and the Law* (1998), 417.

contractual relationship extends into the future, the possible changes that may materially affect the economics of the parties' relationship grow beyond the capacity of contracting parties to predict or manage with tolerable reliability. The costs of contracting rise and the expected benefits grow more risky. Thus an important function of contract law is to supply standard default terms that can free parties from the cost, when they elect to do so, of specifically contracting concerning unknown future risks.

Another goal of the law of contracts is to offer assurance to contracting parties that promises embedded in lawful contracts will be enforced according to law. The topics of performance and breach constitute a large part of the law of contracts and it is of course in this the area that the implied covenant of good faith has its effect. Breaches may occur for a variety of reasons, but the reason that is most important to the law is arguably defaults in performance occasioned by changes in market conditions. Contracts are entered into because at the time of contracting they create rights that are perceived to advance the joint wealth of the parties. But as markets move between the date of the contract and the time of the performance, that which seemed optimal to both parties (ie, was preferred to all available alternatives) may cease to be so seen by one of them. The rules for contractual remedies provide as the general remedy an award of damages measured primarily as the expectational loss of the promisee.[9] Hence if a party does not tender the contracted for performance in breach of its obligation, it will be required to pay to the counter-party damages sufficient to place that party in as good a position as performance would have done.

Thus a futures contract to deliver a commodity in thirty days at a stated price may turn out to be sub-optimal to the seller either because the market price of the commodity rises substantially before the time for performance or because the cost of producing the subject of the contract increases. The question of whether the seller is in breach of the contract is not usually legally problematic. In our example, the contract implicitly but clearly assigns these risks to the seller. Unless a party brings itself within the narrow rules of excuse because of impossibility or frustration,[10] if that party fails to perform it will be required to pay the losses the buyer suffers upon covering. In the event that the buyer's damages are less than the gain that the seller may realise from breach, seller may breach, compensate buyer fully and still profit. Of course such a seller may be constrained by a reputation cost that such action would entail or by a loss to this specific profitable relationship, and thus may not act in increase its immediate proceeds of sale. But putting those things aside, a breach in such circumstances is likely to increase the

[9] American Law Institute, *Restatement of the Law of Contracts, Second* (1981), ch 16.
[10] American Law Institute, *Restatement of the Law of Contracts, Second* (1981), ch 11.

joint wealth of the parties to a greater extent than performance will. This of course is the notion of an efficient breach.[11] Assuming that all expectancy losses of promisee and all consequential damages are paid as a cost of the breach, such a breach is socially desirably, as it increases total wealth. (That is, it is likely to get the subject matter of the contract into the hands of the higher valuing user faster and with fewest transaction costs.[12])

The central institution of expectation damages in our contract law confirms that the law has as one of its purposes the offering of assurance to contracting parties that promises may be reasonably relied upon. These assurances—contractual remedies—however do not really constitute the end of the contractual process, despite the way we present this process in a law school syllabus. As Ronald Coase famously instructed us, legal entitlements provide the background or starting point for bargaining and this is true of contract remedies as well as other forms of legal entitlements.[13] Coase's famous formulation of his theorem was simply that in a world where there were no costs associated with bargaining, the allocation of legal entitlements would not affect the distribution of productive assets at the end of bargaining.[14] Coase's powerful point was that absent barriers of cost or information (search costs) people would re-contract as long as there was an efficiency gain (ie another dollar or euro) to be had.

Everyone well understands that we do not live in a Coasian world of zero transaction costs and symmetrical information. Much of what makes fashioning legal institutions both interesting and difficult comes from the facts of pervasive information asymmetries, positive information costs, restricted cognitive capacity, and opportunistic actors. But even in this actual world the Coasian insight about re-bargaining is important to keep in mind. Re-bargaining between contracting parties will often be a preferred way for commercial parties to resolve differences arising from unanticipated events.[15]

[11] Eg, Kornhauser, above n 9; Goetz & Scott, 'Liquidated Damages, Penalties and a Just Compensation Principle: Some Notes on an Enforcement Model and a Theory of Efficient Breach' (1997) 77 *Colum L Rev* 554.

[12] The law might in this case specifically enforce the contract and then allow the promisee to re-convey at a profit to the higher valuing user but there are two reasons why this might be less desirable from a social perspective. First this would entail additional costs—the cost of the third party finding the promisee. The higher valuing user will have most likely found the promisor and had one negotiation with him. Second, it will take time as well and money to find the promisee and negotiate with it.

[13] Ronald H Coase, 'The Problem of Social Cost' (1960) 3 *JL & Econ.* 1.

[14] Coase didn't suggest that even in the hypothetical world of zero transaction costs that initial endowments would not affect the distribution of wealth at the close of bargaining. Manifestly it would. Thus his point is only about allocation efficiency not about distribution fairness.

[15] See, eg, Frank H Easterbrook and Daniel R Fischel, *The Economic Structure of Corporate Law* (1991), 34 (Coasian analysis applied directly to formulation of policy for corporation law).

Re-bargaining is of course a socially useful activity. It provides a potentially far more flexible and quicker method to adjust contractual disagreements than does adjudication. Thus legal system devices which facilitate effective re-bargaining have value. In order for re-bargaining to occur, however, both parties to a contract must perceive the possibility of a net gain from such bargaining.

The clarity with which legal rights are defined in a contract will have an impact on the desire of the parties to re-bargain. Consider a circumstance in which a risk has been clearly assigned by the contract to one of the parties. Upon the occurrence of this risk, one of the parties at least is faced with a new and presumably unwelcome state of affairs. Its options are as usual either to meet its legal obligation under the new state of affairs or to breach. In the event of breach it will presumably be liable for the payment of expectancy damages plus any consequential damages that may be foreseeable. Re-bargaining of course is a possibility, but given that the parties contemplated the risk and assigned (and priced) it, there is no windfall or unexpected loss in the event's occurrence. Thus the contract's clarity will afford little opportunity for mutual gain on re-bargaining. The promisee is entitled to recovery and the clarity of the rule means it has small reason to compromise its entitlement in any significant way. Thus splitting the transactions cost of collection provide the only fund from which promisor may gain in bargaining.

As the complexity of a contract grows, so too does the possibility that the limited information and cognition of the parties will be supervened by unforeseen events that have consequences to the parties. Because contracts of extended duration are intended to guide the relative legal rights of the parties for a long time, during which variables relevant to the contract may change, such contracts include terms or provisions that are broad or general in nature. The parties can deal in a loose and general way with respect to some areas that are imperfectly foreseeable, but which may recognizably be expected to give rise to future uncertainty. Thus goods under such a contract may be required to be of 'merchantable quality,' or quantities may be set at 'reasonable requirements', or effort defined as 'reasonable best efforts.' But limited human cognition and the diminishing returns on investment in information and negotiation mean that other types of uncertainties, the occurrence of which may create unexpected gains or losses, may not be jointly appreciated at all by the parties.

In circumstances where the parties themselves have not allocated these risks in their contract, absent an implied legal duty to share gains or losses in some way, the occurrence of an unaccounted for risk will potentially confer a windfall gain or an unexpected loss on one of the parties (or unevenly on both). This may strike us as unfair, if we wish to see the contract as a kind of arrangement in which two parties entered a relationship each for its own benefit, but subject to certain unknown risks that are unidentified and unallocated.

It may also be inefficient. A rational party will assume that it has a 50 per cent probability (assuming two parties) of benefiting from, and a 50 per cent probability of being hurt by, an unanticipated risk. Risk averse individuals or institutions will demand compensation to bear this risk but will be unable to quantify it well. Thus a truly unanticipated event is precisely the sort of risk that the parties could most effectively negotiate about *after the risk of it is known* (which in some instances will not occur until the event the risk is associated with has itself occurred). But the dilemma is that *if* the legal default itself somehow defines the risk *with clarity* (by, for example, privileging a default rule that holds that a party makes no covenant that is not an express covenant or one clearly implied by what is express), then re-negotiation will not be an effective way for the parties to deal with unacknowledged risk once it is anticipated.

One response would be for the legal systems to create some degree of ambiguity in order to facilitate re-negotiation after the revelation of an unanticipated risk. It is useful that re-bargaining occurs after an unanticipated event occurs. It is only then that the parties have information that allows them to allocate its impact efficiently. If a clear default rule allocates the impact of a non-foreseen event, there will be little room left for re-negotiation.

a. The Implied Covenant in Action

When will it be socially advantageous for a court to imply this duty of commercial good faith and what does it require when it is found to exist? What are the probable costs of the existence of this duty, and how may they be contained or eliminated? After some general discussion we will return to our hypothetical.

Legal Tests If, from a social perspective, we are concerned to constitute contract law as a set of rules, principles and practices that assists in the creation of wealth, we must of course be concerned that any innovation in that body of law does not do more harm than good, ie does not impose greater costs than benefits in wealth creation. Admittedly a doctrine that permits courts to address the parties' relations from an *ex post* perspective and enforce a duty as inherently ambiguous as one of good faith, may easily sow the seeds of doubt and concern among contracting parties *ex ante*. Such doubt is yet another source of risk, which if unhedged will have costly effects. Therefore we suppose that the first task with respect to the doctrine of commercial good faith is to attempt to construct a legal test for its use that will contain or restrain unwanted risk. So we turn to the legal test that we propose that courts should employ when a party is charged with violating an implied duty of fair dealing and good faith.

The first and nearly universally accepted rule is that an implied duty must be construed to be consistent with the express terms of the contract.[16] This rule is an important protection against the costly ambiguity that might exist were courts encouraged to re-write the parties' allocations of risks in a written contract. Where the parties themselves expressly address a topic, absent deception, mistake that justifies relief, or other ground for relief, their allocation of rights and duties must be respected if the fundamental utility of the institution is to be protected. This principle constrains the risk of costly ambiguity that judicial construction *ex post* of unstated duties may create. In our view this foundation limitation on judicial action is primary.

Second, assuming that a fair reading of the rights and obligations created in the contract does not clearly provide an answer to a dispute arising from contract, one must ask whether either party is under a duty that is unexpressed in the contract, but is consistent with the express agreements that have been made. In answering this question courts tend to fall into one or two conceptual camps. The first approach looks to the traditional doctrine of contract law. For those who occupy this position the fact that a court is implying an *agreement* forces them to focus imaginatively on the moment of contracting. They appreciate that the parties must have understood that they did not perceive all of the forces that might and would over time work upon the relevant environment for their agreement. The parties likely created some protections against unforeseen developments (a contract term or a termination right, a process for re-pricing, etc), but in this view the parties knew also that they did not have perfect information. Some risks were assigned in the contract but others were unconsidered. Given a complex environment into which a contractual relation is inserted, it is almost inevitable that unforeseen contingencies may arise. One simple explanation of this fact is that at some point the cost of writing additional terms outweighs the benefit.[17] That is, as fewer and fewer probable events are considered, costs rise and likely benefits fall.

Courts that respond in this traditional way to contract disputes arising from unallocated risk place themselves imaginatively in the place of the contracting parties at the moment of contract. They then tend to enforce the terms that the parties 'would have wanted'[18] had they thought to contract about it. Support for this approach can be said to derive from the contractual nature of the parties' undertakings. Thus a court interpreting the meaning of their acts is well grounded when it enforces obligations that (the court concludes) the parties themselves would have agreed upon, had they had the information necessary to contract at all on the topic.[19]

[16] Eg, *Katz v Oak Industries, Inc.,* 508 A. 2d. 873 (Del Ch 1986).
[17] Ayres & Gertner, 'Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules' (1989) 99 *Yale L J* 87, 127.
[18] *Ibid.*
[19] See *Katz v Oak Industries*, above n 16.

In this view, therefore, the measure of what good faith requires is found from the parties themselves.[20] They are required to perform the contract according to the reasonable expectations of (hypothetical) fully informed persons in their position.

In *Katz v Oak Industries, Inc*,[21] the court recognized that 'the parties occasionally have understandings or expectations that were so fundamental that they did not need to negotiate about those expectations. *Corbin on Contracts,* (Kaufman Supp.1984) § 570.'[22] The court also acknowledged that modern contract law tends to enforce these unstated, reasonable expectations by employing the language of implied covenants of good faith.

The court stated the test it would apply as follows:

> Is [it] clear from what was expressly agreed upon that the parties who negotiated the express terms of the contract would have agreed to proscribe the act later complained of as a breach of the implied covenant of good faith—had they thought to negotiate with respect to that matter. If the answer to this question is yes, then … a court is justified in concluding that such an act constitutes a breach of the implied covenant of good faith.[23]

Of course, such questions will frequently be quite difficult to answer with confidence. But sometimes this will not be the case. When the purpose of the parties' bargain is discernable and the nature of the overlooked eventuality is such that a court may imaginatively envision with confidence how the parties would have dealt with this unexpected effect, then the court is free to fashion an implied legal duty. In all events under this approach a determination of the constructive (or imaginary) intentions of the parties regarding their performance under unforeseen circumstances will be based on all of the particularities of the case, including the type of contract, the parties' goals in entering the agreement and all other circumstances.[24] In this way the construction of the duty of good faith in commercial contracts is quite like the construction of the contours of the duty of loyalty in US corporation law, albeit that the contract itself may constrain that interpretative enterprise more than does the corporate statute.[25]

A second view exists as to how courts might deduce breach of an implied duty. This perspective is premised on the assertion that there is no such thing as constructive intent of the parties when they had no actual intent. This second camp of commentators and judges asserts that it is too speculative

[20] Jeannie Marie Paterson, 'Good Faith in Commercial Contracts? A Franchising Case Study' 29 *Australian Business Law Review* 270, 277.
[21] See *Katz v Oak Industries*, above n 16.
[22] *Ibid*.
[23] *Ibid*.
[24] Jeanne Marie Paterson, above n 20.
[25] See William Allen and Reinier Kraakman, *Commentary and Cases on Corporate Law*, (Aspen Law & Business Press, 2003) ch 9.

to construct after the fact a view of the counter-factual intention at the time of contracting. In trying to answer the question of whether an implied covenant has been breached or not, this cohort of judges and commentators prefers to turn immediately to standards of fairness and justice (inferred from the body of relevant law) while searching for the interpretation which maximizes the expected value of the transaction.[26] This approach is sometimes described as the 'communitarian' approach, since 'reference might be made to [standards of behavior, practices and ethics of] the business or industry in question'.[27] Under the Uniform Commercial Code, good faith is said to mean 'honesty in fact and the observance of reasonable commercial standards of fair dealing *in the trade*'[28] (italics added). The US courts, therefore, will look to the substance of the agreement, while adopting the standards of the relevant industry, or 'idealistic concepts of the community' if the standards of the relevant industry are insufficiently high.[29]

Each of these approaches provides a rough guide to a court in working out a particular problem of contract failure. We suggest that the approach that focuses on the contractual nature of the implied obligation to be the preferred approach, for the reasons we now briefly set forth.

The essential thing that separates contractual undertakings from many other obligations imposed upon us by tort law, criminal law, or regulatory statutes, is the *voluntary* acceptance of a duty. Among the most beneficial aspects of the voluntary nature of contract as a means to organize social action is its flexibility. Two persons—the smallest unit of co-ordinated social action—may custom tailor their rights and obligations to each other, in light of the known particularities of their environment and the particular objectives that each seeks to achieve. As the popular song of a few years ago put it, 'nobody does it better.' And the reason no one is likely to structure the commercial interactions of two persons better is that no one knows better than the parties themselves their aims, and no one has better incentives than they to uncover relevant information bearing on the performances that they contemplate. Thus respecting the power of parties to shape the precise nature of their undertakings is one of the basic building blocks of a decentralized legal system in which human liberty is a foundation and contract a fundamental tool of achieving market efficiency.[30]

The contractualist approach to constructing a specific meaning to an implied covenant of good faith is more consistent in our view with the

[26] *Farnsworth on Contracts*, (Boston, Little Brown & Co, 1990), pp 500–501.
[27] Jeannie Marie Paterson, above n 20, 278.
[28] *Uniform Commercial Code*, s 2–103(1)(b).
[29] Jeannie Marie Paterson, above n 20, 278.
[30] R Posner, *Economic Analysis of Law*, 4th ed (Boston, Little Brown. 1992), 13. The use of 'market efficiency' here refers to the concept of Pareto efficiency, ie, none of the contracting parties believes he is worse off and at least one of them believes he is better off as a result of entering the transaction.

basic structure of contract law than is the communitarian approach, in which a court imposes its views of justice and fairness on the parties. The contractualist test seems to us less likely to impose results inconsistent with the allocation of risks and benefits that the parties themselves expected, and more likely to provide a benefit. Resort to a perceived general standards in the community, including a more particularised community (or trade) in which the contract was negotiated, invites the logical possibility of a result different from the result that the parties themselves would have elected had they had full information at the time of contracting. Results that deviate from what the fully informed parties would have preferred can of course be justified, but not by reference to the parties' own goals. Such results can be justified by reference to a social (legal) preference that, while costly to the individual parties, furthers a different social goal than the parties' perception of their joint welfare. Certainly such social goals exist and justify outcomes other than the two parties might intend. (Refusal to enforce meretricious contracts—eg prostitution—is a common example of a trump over individual choice.) But when courts construe contracts that are not themselves unenforceable, it seems odd that the legitimate exercise of social power should trump individual choice. As a practical matter, since society (or relevant sub-units or communities within it) holds complex and at some level inconsistent values, deployment of the broader communitarian test subjects the parties' presumed or constructed joint wealth-enhancing objectives to possible trump by a judicial vision of community standards.[31]

Implied Covenants in the Re-Negotiation Process Thus we prefer the approach to interpreting implied covenant of good faith of those courts that attempt to construct a vision of what fully informed contractors would have chosen had they known all of the facts as they have evolved. But in our view the utility of this doctrine is less importantly bound up with which of these two legal tests is deployed in court interpretations than it is with how this doctrine affects the parties' interactions after the revelation of an unexpected development. It is in this setting—re-negotiation—that we believe that the legal concept creates social value.

Let us turn to a consideration of our hypothetical case to try to show how the presence of an ambiguous requirement to treat contractual partners in good faith may affect the resolution of this situation. Later we will try to persuade you that the resolution that we would predict with the implied covenant will be not only intuitively more attractive from a fairness perspective than the outcome we would expect without this ambiguous term, but that it will be more efficient *ex ante*.

(a) The comparative baseline: No implied duty: Absent the implied obligation of good faith and fair dealing, existing case law in the US suggests that

[31] See also Jeannie Marie Paterson, above n 20, 279.

Bigco may be permitted to close the Texas plant as it did not contractually bind itself to operate the plant for the period. It did not agree to restrict its inherent right to close down the plant or transfer the production. This result, absent an implied duty, would perhaps be expressed in assumption of risk terminology. If Supplier wanted a covenant to keep the plant open for the entire period , it should have obtained a covenant.

This result would reflect a default rule of contract interpretation to the effect that benefits or detriments falling upon the parties to a contract by reason of events not contemplated in the contract remain where fate (or better, future events) deposits them. This is a clear rule (relatively at least): if a party does not explicitly or by clear implication restrict a counter-party in the exercise of any right possessed by that counter-party, then there is no contractual restriction.

This rule would not help the parties adjust the terms of their relationship in light of an unexpected factual development. What do we imagine that the parties themselves would have wanted had they had the prescience at the time of contracting to imagine that a breakthrough technology might occur (but might not) during the eight year term that they were considering? Might they have negotiated that Bigco's buying obligations would remain totally unaffected by this fact (ie, all of the gains that the technology might create would be foregone by Bigco)? Might they instead have agreed that in that event Bigco would be free to take advantage of any gains by moving its operations to a different plant (ie, all of the losses to be borne by Supplier)? Either outcome is conceivable and if either had been bargained we could assume that the contract benefits to each side reflected the expected costs and benefits of that bargain. But in the event that the parties did not address this risk, it seems wooden and unhelpful for the law to reflect a default term that chooses either of these extremes.

Moreover, what if the parties re-bargain where Bigco has a clear legal right to cause this result (that is, to impose these losses on Supplier)? What leverage does Supplier have in this world? It wants to reduce its losses, and rationally it would transfer to Bigco so much of the value of its contract as is necessary to preserve some value. That is, Supplier will be forced to sell at some price just above $1.60 a pound (how much above would depend upon its estimate of Bigco's cost of shutting the Texas plant and shifting production elsewhere). If the default rule is the opposite, then Bigco will either have to pay the contract rate or pay all of Supplier's damages. That is, the losses will fall principally on Bigco.

In the event that parties in fact fail to negotiate about a future albeit unexpected event, do we really imagine that a result where all of the gains or all of the losses are allocated to one party is what they would chose?

(b) The implied good faith alternative: Consider now the outcome of renegotiations in a legal system that recognizes an ambiguous obligation on

each party to treat its counter-party in good faith. In this system Supplier is supplied with an argument that Bigco's threatened removal of its operations constitutes a violation of the default good faith obligation. How plausible this argument can be made to seem will affect the results of any re-bargaining. For example, Supplier might plausibly argue that so long as Bigco is making a profit it should be regarded as obliged to continue operations during the course of the contract term to which the parties agreed. No external fact, it would say, has made any other step essential. As Bigco is commercially able to continue as the parties had contemplated (ie, it is making a profit), these arguments are not frivolous in a regime that would require commercial good faith. Would they, or should they, prevail?

We need not express a view on this subject. It is sufficient for our purposes to show that with this additional leverage the outcome of the re-negotiation will be less exploitative than under the clear rule standard. The lack of clarity in the legal requirement will aid the commercial parties in adjusting to the new reality in a way that may preserve the relationship. One respected jurist has noted this fact nicely:[32]

> It is a nice question how exigent the buyer's change of circumstances must be to allow him to scale down his requirements ... Yet, although more than whim is required ... how much more is unclear. There is remarkably little authority on the question. This is a good sign; it suggests that, while we might think it unsatisfactory for the law to be unclear on so fundamental a question, the people affected by the law are able to live with the lack of certainty. The reason may be that parties linked in an ongoing relationship—the usual situation under a requirements contract—have a strong incentive to work out disagreements amicably rather than see the relationship destroyed by litigation.

An implied obligation of good faith facilitates such working out.

B. Efficiency Implications of the Implied Obligation of Good Faith

It may strike one that the legal system is more likely to reach a *fairer* result if we afford to the Suppliers of the world some leverage in such situations. One response to such an intuition may be, 'Why should we care?' If an unforeseen event occurs (eg the change in technology here) it may produce gains or losses that will somehow fall upon the parties. Why should society care how they fall, if the parties have not contracted with respect to such risk? For some law and economics scholars, 'distributional' consequences of rules—that is the distribution among parties of losses or gains—seem decidedly secondary. These scholars focus their attention on the allocative

[32] *Empire Gas Corp v American Bakeries Co* 840 F2d 133 (US App 1988).

effects of legal rules, focusing in particular on the question whether certain rules tend systematically to produce more or less total wealth. But in most cases judges are neither trained nor oriented to create rules that aim to produce increased total social wealth. They are often confronted with questions of distributional fairness. It would be quite impossible to present a descriptive account of how courts behave to exclude distributional or fairness concerns. Our legal system *is* concerned with allocative efficiency, we believe, but generally in a cumulative way. In individual cases courts are often affected by fairness or distributional concerns. Judges, like most others in the society, tend to take cognisance of the distributional effects of their rulings.

Thus, in the re-bargaining process, it no doubt strikes many of us that results arising from a bargaining process in which each party has some plausible leverage over the other are likely to be ethically superior (fairer) than those results imposed upon one party by another. Thus the facilitation of a re-bargaining process by recognising an implied ambiguous term of a contract has, we suggest, an ethical appeal on distributional grounds. But we do not base our approval of the implied term of good faith and fair dealing on this ground alone. Rather we suggest that the implication of such a term may also be viewed *ex ante* as assisting contracting parties in reaching agreements and thus is likely to promote efficient outcomes as well.

Obviously someone, like Supplier, who makes substantial capital investments that cannot easily be deployed in an alternative use may be at risk of opportunistic exploitation by their counter-party. (By contractual opportunism, we generally refer to action benefiting one party and injuring the economic interests of the counter-party, which action appears to be legally permissible, but is inconsistent with the parties' unexpressed shared assumptions at the time of contract formation). Once the investment is made by one party, a counter-party understanding the loss that would be entailed were that party forced to the alternative use of the capital, can attempt a re-bargain that extracts value from the party who has sunk costs. Of course, contracts should be designed so as to minimize this possibility, but contract writers will be fallible in foreseeing all such opportunities. Oliver Williamson uses the fact of asset specificity—investment in assets with much lower second-best deployments—as a key explanation of why we have seen shared ownership structures (including corporations) evolve.[33] In contracts in which the parties do not solve opportunistic re-bargaining through common ownership, other devices need to be employed. We suggest that the implied obligation of good faith may be seen as a default term implied by law to help serve this purpose.

[33] See Oliver Williamson, *Economic Insitutions of Capitalism*, (Harvard University Press, 1985), ch 4.

As a default rule that addresses circumstances that were not in fact contemplated in the express terms of the contract fairly interpreted, the implied duty acts as a sort of partial insurance (or risk sharing between the parties), which reduces the risk of losses associated with the unspecified event, including opportunism. The 'insurance', has a cost. Parties to contracts will, to some extent, give up the prospect of benefiting unilaterally from an unexpected event that in fact benefits them disproportionately. Thus a risk-preferring contracting party might not choose to have an implied duty of good faith in her contract; she might prefer to see how unanticipated events work out. Presumably two sophisticated parties could agree that they prefer to roll the dice on the future, and could rationally elect to exclude this implied covenant. (They may not do so in contracts for goods under US statutory law, as noted above, but we take that as indicating that the law is designed for consumer transactions and that the parties there are not thought to be sophisticated or indeed even to engage in actual bargaining.) In fact, however, few persons, including sophisticated parties, would prefer to accept unknown and thus unquantifiable risks if this could be reasonably avoided. Even those who enjoy vacationing in Las Vegas enjoy taking risks about which they can make some calculations. Most of us are risk averse, especially when it comes to bearing financial risk. If we are not compensated for bearing financial risk, we prefer not to do it.

Thus, the implied obligation of good faith can be seen as a way to moderate risk that is beyond the informational capacity of the parties—or which is uneconomical for parties to attempt to estimate and value *ex ante*. There will always be such risks, and creating default protections against them will assist contractors *ex ante*.

We have now explained how we believe the implied term works as partial insurance and to soften counter-party opportunism. It will create ambiguity concerning what legal responsibility may require of the party upon whom fortune has smiled or, as in the case of our hypothetical, who has arguably acted opportunistically. In either event the unsettled nature of what the law may require will improve the situation of the party upon whom fortune has frowned (or, as in our hypothetical, from whom value is sought to be extracted).

Thus we conclude that the availability of this form of protection will, at the margin, allow contracting parties to enter into contracts more freely. It is a rational part of a legal structure governing contract-making that recognizes that contracting parties will inevitably only be able to foresee imperfectly the environments in which their contractual performances will be called forth. If we assume that contracting parties are generally risk averse, and that facilitating the making of investments in long-term economic relations is a valuable goal of contract law, then one can appreciate, we think, that the ambiguity inherent in an obligation as vague as one to exercise commercial good faith can itself add value.

Commentary on 'The Uses of Ambiguity in Commercial Contracts: On Facilitating Re-Bargaining'

PAUL LOMAS

THIS PROVOCATIVE CHAPTER by Professor Allen and Ms Levy focuses on the economic value of including 'good faith' obligations in contracts; a deliberate policy of introducing ambiguity, of incentivising re-bargaining. The seminar at which it is presented, mixing academic thinking, judicial experience and the comments of practitioners, provides a wonderful opportunity for a review of a topic so open and ripe for development as the potential role for good faith in English contract law.

Moreover, the topic is highly appropriate. Sir Christopher Staughton made the telling point in an article in the Cambridge Law Journal in 1999,[1] that academic institutions were exceptional in teaching contract law, but devoted no time to the critical issue of interpretation. The issue, in fact, bedevils practice on a continuing basis. Lord Steyn makes the point very strongly, from the judicial viewpoint, in his own paper at this seminar; and he only sees the cases where it goes wrong!

However, the question posed by Professor Allen's admirable and thought-provoking paper is not really contractual interpretation itself. This is fortunate.

It enables me to steer clear of the delicate debate on the correct canons of interpretation of contracts as developed recently by the courts, including the House of Lords, on various occasions. I can, for example, side-step the issue of whether *ICS v West Bromwich Building Society*,[2] *Mannai Investments v Eagle Star*,[3] or *National Bank of Sharjah v Dellborg*[4] represents the correct

[1] Sir Christopher Staughton, 'How Do the Courts Interpret Commercial Contracts?' 58(2) CLJ 303.
[2] [1998] 1 WLR 896.
[3] [1997] AC 749.
[4] Unreported.

approach to the relationship between what the reasonable man would have intended and what the plain words say (against the background of the omnipresent factual matrix which we are to investigate with an intensity which shows distinct signs of variability!).

Rather, the question posed by the paper is whether it would be beneficial, in social and economic terms, for English law to adopt the trend seen in the US, in some Commonwealth jurisdictions and often in civil law jurisdictions, towards parties being subject to a general duty of good faith in relation to the way in which they fulfil their contractual obligations.

Professor Allen, elegantly, and delicately, suggests that it would. His central thesis is that it is or can be beneficial to society, in economic terms, to do so—'the implied doctrine of good faith is useful in long-term contracts precisely because it is ambiguous'. In particular, he suggests that this is so because, by creating constraints on behaviour within a contractual situation (those constraints arising from the fact that the risk trade-offs associated with litigation are adjusted by the duty to behave with good faith) parties will be encouraged to renegotiate, to re-bargain to reach socially desirable outcomes, rather than to enforce arbitrary and not necessarily efficient outcomes.

This is particularly relevant in the context of long-term contracts, especially in the area of supply or energy contracts. This is so because the parties cannot foresee the various pitfalls that lie ahead, and even if they could or attempted to provide for the range of possible outcomes, the transaction costs in addressing them would increase dramatically, no value being thereby added since most of these outcomes would not occur.

How much simpler therefore to have good faith to ride on a white charger over the hill to rescue one unfortunate party from his bargain when the market moves against him and his counterparty enforces his rights. This is seductive thinking, particularly when made sensuous by its intellectual foundation on economic theory and the delivery of social good, of greater wealth, and of enhanced economic strength.

Professor Allen analyses the basis of the operation of the duty on two grounds. The first is the contractual basis of seeking to determine what the parties would fairly have agreed. The second is the so-called 'community basis' of imposing society's values of what it regards as fair. He plumps for the first. This must be right in terms of giving the balance of credit in a free society to the rights that free individuals have freely chosen to adopt. This is the less radical and more conventional approach. I have no quarrel with this—if one must choose between these two options.

Divertingly, Professor Allen suggests that in a Darwinian model of competing legal systems, the most economically efficient system will win out. Fortunately, it is not his point that this is the reason for the strength of the US economy—which might open a rather wider debate! Rather the concept is that the most efficient legal system will flourish. This is an inherently attractive idea; I question, however, whether there is the necessary mechanism, even capitalism—red in tooth and claw—for this evolution to operate. There has been no obvious stampede by commercial parties to jurisdictions offering the good-faith option on their menu.

However, I welcome the principle of addressing these issues through the lens of economics. This is wholly refreshing, even though it forces me to accept again what I long ago realised: that I was merely a transaction cost, and 'indeed' have a vested interest in remaining a transaction cost.

In keeping with that limited role, and given my background in commercial litigation, it is, however, impossible for me to resist challenging Professor Allen's thesis. I would therefore venture to set up a counter thesis:

1. English law has set its face against importing such a principle;
2. it has been right to do so; and
3. English law permits the parties to choose to import equivalent duties to those of good faith if they deem it necessary—this is the true economic freedom.

I illustrate this by reference to some of the details of Professor Allen's chapter.

As he admits, ambiguity is costly. One of the difficulties that I have with the paper is that his example (Bigco purchasing the increasingly prevalent and pervasive 'stuff' from Supplier) fails to take account of this risk and its associated cost. In an effective market, the risk of Bigco terminating should have been priced into the price offered. Conversely, the risk for both parties inherent in a general good faith obligation should increase the margin that both parties need from the transaction—in theory, this increased margin could result in transactions not occurring that, from an economic net benefit point of view, should occur.

The chapter meets this in part by accepting that the doctrine cannot override an express term. However, with respect, although that may manage down some of the risk, it does not remove the problem.

It is also important to note that a different assessment of risk creates value (indeed it drives a market). If party A assesses a risk of a given event as low and party B assesses it as high, both parties can trade that assessment of risk by giving protection within the contract which enhances overall value: it matters not that the event, ex post, either will or will not have occurred; *ex ante,* the acceptance or laying off of that risk will create value at that time. However, this will not occur if there is some overlay of good faith which removes 'unfairness' after the event, but only by, at the same time, extinguishing the ability to manage that risk.

It is also important to distinguish between (i) using good faith to fill a gap—where an event happens, short of frustration or force majeure for which the contract is simply silent—and (ii) the case where the contract does not protect a party against a particular event which exposes (clearly and unambiguously) that party to costs which, with hindsight, seem unfair (even though the risk of such an occurrence, if we are content to live in an economist's classical world for a moment, should have been priced into the contract).

The first is clearly catered for in English law. We are now in the classic territory of interpretation, one might say interpolation. The traditional approach is well established: objective intention (leaving aside what that means, for a moment) against a factual matrix (leaving aside what goes in that for another moment), market practice and custom, a sensible result for which the parties could have contended. Thus far, English law does not have a problem. Indeed, thanks to the combined efforts of many of the country's finest judicial minds over the past few years, Professor Allen's economic benefit theory, derived from uncertainty leading to a re-bargained outcome, is well established in the English legal sphere!

Whilst Professor Allen does not rigorously maintain a distinction between the two roles of good faith, it seems, in particular from his example (where there is no ambiguity), that he intends to address the second: 'good faith and fair dealing' as an implied term, as set out in the Restatement.[5]

In part, he draws comfort from the Chancellor, or his foot, from duties arising in Equity. However, English courts have long eschewed that temptation. Indeed, a classic English textbook on contractual interpretation (Lewison[6]) has not one entry in the index under 'good' or indeed 'faith'.

Save in very specific areas (one thinks of *Quistclose* trusts,[7] agency relationships, confidential information and certain forms of capital market instruments, which are in fact wholly consistent with the principles of equity), parties have the freedom to contract and the responsibility that intrinsically and irrevocably accompanies it. These equitable duties are reserved for those whose interests need protecting and do not stray into, for example, long-term commercial energy contracts between consenting grown-up corporations.

Moreover, although he throws himself on economic efficiency as a touchstone, Professor Allen must address two issues which trouble me about his analysis.

First, it is clear that judges have proceeded in the direction that he advocates not because of a deep familiarity with the theorems of Coase on rebargaining[8] or Kaldor/Hicks on compensation[9] but on the basis of what they thought was fair between the parties—thus the economic outcomes were either deeply intuitive on behalf of the judiciary or wholly serendipitous.

Secondly, and here I am taking a certain degree of personal risk and placing an excessive degree of weight on my own limited knowledge of economics,

[5] American Law Institute, *Restatement of the Law of Contracts, Second*, § 295 (1981).
[6] Lewison, *The Interpretation of Contracts*, 2nd edn, (London, 1997).
[7] *Barclays Bank Ltd v Quistclose Investments, Ltd* [1970] AC 567.
[8] Coase, R H, 'The Problem of Social Costs' *Journal of Law and Economics* (1960), v 3, pp 1–44.
[9] Kaldor, N, 'Welfare propositions and interpersonal comparisons of utility' (1939) *Economic Journal*, 49:542–549 and Hicks, JR, 'The foundations of welfare economics', (1939) *Economic Journal*, 49: 696–710.

surely what increases wealth in society is an increase in the productivity and the economic efficiency of the relationships between the parties. They should adopt Pareto's theory and move progressively towards the Pareto efficient limit where no further economic efficiency can be achieved under those boundary conditions. However, a considerable part of the paper is concerned with distributive effects and the allocation of benefits between parties. A fairer system may well be more attractive in social terms, it may reflect our concept of what is just; but it does not obviously or directly enhance productivity.

I could accept Professor Allen's proposition that legal devices facilitating rebargaining in principle have a value (even though this may not off-set the losses that arise elsewhere in the system). My point, however, is that you do not need to import a doctrine of good faith to achieve this. Litigation has that element of risk already. Every practising lawyer at this seminar is familiar with the 'litigation risk' discussion with a client.

Moreover, parties can expressly provide for the incorporation of a good faith mechanism. My issue is not with the concept of good faith. In fact I am rather in favour of it. Rather like Mahatma Ghandi's comment on western civilisation, I think that it would be an excellent development. Rather it is with the mandatory imposition of a duty of good faith in the context of the conduct of stipulated contractual relations that I have an issue.

Lord Ackner in *Walford v Miles*[10] famously remarked (but I paraphrase) that a duty to 'negotiate' in 'good faith' is an oxymoron; in his view good faith is inherently repugnant to the adversarial position. However, many jurisdictions do not take this view and the dictum has been criticised on the basis that, if best endeavours and other subjective standards can be monitored by the courts, why is it that good faith is inherently different; indeed why is it that good faith in an adversarial position is different?

In reality the difficulty with such clauses is the absence of certainty as to a result, not the imposition of good faith on the procedure.

Provided that the parties have avoided entering into an agreement to agree and the contract is not too uncertain, which usually means that there are clear standards, and that there is an exit route from the process (for example an expert clause or a pendulum arbitration) such difficulties can be handled. Such processes are supported by the Courts. It is trite that expert determination clauses are valid. In October 2002, in a case in which I was involved, Coleman J upheld a contractually stipulated mediation clause that provides some of the rebargaining incentives that Professor Allen seeks.

It is not surprising that many of the cases displaying a 'strict' interpretation of the wording of a contract in English law have been in the energy field, where long-term contracts have been put under strain by the volatility of

[10] [1992] 2 AC 128.

energy prices and parties have operated and enforced their strict legal rights in ways that suited their classical economic theory/utility maximising interests. Cases like *Total v Arco*[11] (House of Lords) and the J-block litigation (see judgment of Potter J in *Phillips v Enron*[12]) show that parties will be held to the bargain. But let us not forget that these are large commercial entities managing risk in volatile markets. It is far from clear why the imposition of a blanket duty of good faith leads to either fairer or more economically beneficial results. This is for two reasons which apply perfectly effectively against the background of the English system of strict interpretation:

1. Parties have resorted to considerable ingenuity to create legal arguments that carry risk for both sides, in order to cause just the level of instability which Professor Allen refers to and to force a degree of re-bargaining.
2. Where the parties have identified an area of risk and the transaction costs associated with analysing it and allocating it are excessive (as they often will be), they can choose softer dispute resolution mechanisms, such as those I have just mentioned—this is the free market *pacta sunt servanda* solution—and it works.

A classic recent example is the history of poor little RETA who evolved into the giant killing NETA. The reform of the market structure for the England and Wales electricity market fundamentally changed the way in which energy prices were set in the UK (with dramatic consequences). Many parties had long-term supply contracts at very large values. In addition, parties had hedged the market or pool price volatility risk using the contract for difference market. These CfDs were intrinsically pegged to a pool-purchasing price which was abolished by NETA.

The CfDs, some of which were quite long term, classically had a clause dealing with abolition of the electricity pooling system since a form of reform was partially foreseen. This imposed a form of renegotiation, in order to establish a difference reference index, with reference to the Electricity Industry Arbitrations system (which has express expert determination jurisdiction) if this bilateral process failed. Since the CfD's were, by definition, out of the market, there was the possibility of a torrent of litigation. However, these clauses worked and the relationships were renegotiated either to a fresh index or to a commuted result. This is a perfect free market solution to an identified, but difficult to quantify, risk.

Thus, whilst I welcome Professor Allen's analysis, I am not persuaded that English law needs to follow this route. Many, if not all, of the benefits that he seeks are achievable today, if by slightly different legal routes;

[11] *Total Gas Marketing Ltd v ARCO British Ltd* [1998] 2 Lloyd's Rep 209.
[12] *Phillips Petroleum Company UK Ltd & Ors v Enron Europe Ltd*, 1996, unreported.

I worry that the costs of the uncertainty that he accepts will outweigh the marginal benefits (if any) that might be obtained.

However, (i) accepting, as I do, that I am but a mere transaction cost; (ii) noting the prevalence of litigation in the US market; and (iii) recognising the value that this brings to the legal economy (albeit that this benefit is not something that society at large always values highly), I wonder if it would not be in my own interest to accept his thesis. Maybe that is precisely why it should be rejected!

8

Objectivity and Committed Contextualism in Interpretation

HUGH COLLINS

THIS CHAPTER WAS provoked by Lord Steyn's ambitious and multifaceted chapter about interpretation.[1] His discussion ranges over the interpretation of contracts, statutory interpretation, and constitutional adjudication involving human rights. He suggests some general propositions applicable to all these instances. In so doing, he contends that there are common principles applicable to the interpretation of all legal documents. He quotes with approval Lord Blackburn's remark:

> I shall … state, as precisely as I can, what I understand from the decided cases to be the principles on which the Courts of Law act in construing instruments in writing; and a statute is an instrument in writing. In all cases the object is to see what is the intention expressed by the words used.[2]

On reading this remark, it struck me as an odd thing for Lord Blackburn to say. After all, he was the author of the famous 'objective test' in English contract law, which every law student learns.

> If, whatever a man's real intention may be, he so conducts himself that a reasonable man would believe that he was assenting to the terms proposed by the other party, and that other party upon that belief enters into the contract with him, the man thus conducting himself would be equally bound as if he had intended to agree to the other party's terms.[3]

It is true that this latter statement was pronounced not in connection with interpretation of a contract, but for the purpose of rejecting an argument that if the parties were not in their intentions *ad idem,* there could be no contract. Even so, the objective approach is applied to interpretation of contracts,

[1] J Steyn, 'The Intractable Problem of the Interpretation of Legal Texts', above ch 5.
[2] *River Wear Commissioners v Adamson* (1877) 2 App Cas 743, at 763.
[3] *Smith v Hughes* (1871) LR 6 597, QB.

because by the same logic, if the actual intention of the parties, as opposed to the terms of their agreement, is irrelevant to the formation of a contract, the actual or subjective intention should be equally irrelevant to determinations of what the parties actually agreed. A promisor cannot enforce the terms of the contract in the sense that she may have intended, if the promisee reasonably understood them to contain a different meaning.[4] As Justice Holmes once wrote, 'Nothing is more certain than that the parties may be bound by a contract to things which neither of them intended ...'.[5]

If the objective approach is used for the interpretation of contracts, it seems to rule out references to the actual intentions of the parties. How could Lord Blackburn insist, without contradicting himself, that a search for intention was always relevant? Does the answer lie in the point that his precise expression is that the court should discover the 'intention expressed in the words'? Is the intention expressed in the words somehow not really actual intention at all? Or, to ask the same question another way, what difference does it make to confine the search for intention to an examination of the words used?

The puzzle presented by Lord Blackburn's remark in truth merely presents a perennial uncertainty in the common law of contract. What is the relation between the 'objective approach' of the common law and the idea that contracts are based upon the 'will' or the consent of the parties. In particular, how can the objective approach be reconciled with the frequent statement by courts that their objective in interpreting contracts is to discover the common intention of the parties? These puzzles are only heightened by further difficulties resulting from the ambition of Lords Steyn and Blackburn to generalise their approach to all instances of legal interpretation.

Although the confines of a short essay prevent me from attempting to consider all aspects of legal interpretation, I will develop a few central propositions. First, I consider the implications of the approach to interpretation, which I think Lords Blackburn and Steyn share, that can be described as 'committed contextualism'. I argue that 'committed contextualism' necessarily leads to two further positions: (a) that the literal meaning of a text only matters to the extent determined by the context; and (b) that there can be no unified approach to legal interpretation (other than being a 'committed contextualist'). To explain and support those propositions I reconsider many of the examples discussed by Lord Steyn, but I also investigate in particular, using Lord Hoffmann's judgment in the *Investors' Compensation* case,[6] the implications of an objective approach to the interpretation of contracts. Two briefer sections consider the meaning and merits of a 'purposive'

[4] *Thake v Maurice* [1986] QB 644, CA; G Treitel, *The Law of Contract*, 10th edn (London: Sweet & Maxwell, 1999) 174.
[5] OW Holmes, 'The Path of the Law' (1897) 10 *Harvard Law Review* 457.
[6] *Investors Compensation Scheme v West Bromwich Building Society* [1998] 1 WLR 896, HL.

approach to interpretation, and, what I shall argue is a rather different activity, the application of constitutional human rights documents. My conclusions will be, if I can be so bold as to put them in personal terms, that Lord Blackburn did contradict himself, that Lord Hoffmann was right about contract interpretation, though not perhaps for the same reasons as he thought, and that Lord Steyn, though perhaps not appreciating the full implications of his position of being a 'committed contextualist', nevertheless usually applies this position accurately in his discussion.

I. COMMITTED CONTEXTUALISM

Although Lord Steyn does not use this phrase, throughout his chapter he is anxious to insist upon the need for a 'contextual approach' to the legal interpretation of documents. In the way he presents this position, it seems to have two elements. The first is that the correct approach is to try to discover the meaning of the written text, and not to speculate more broadly on what would have been a sensible thing for the parties to a commercial contract or Parliament to have done—in short: text matters.[7] But Lord Steyn's second contention creates a problem. The meaning of words, he insists, always depends on context, so that the meaning of a text can only be understood by reference to its context.[8] He adds, however, that the contextual approach may leave two or more possible or 'feasible' meanings, at which point consequentialist or policy or purposive arguments must be taken into account to isolate a preferred meaning. His second contention lives in tension with the first. Text matters, but text cannot be understood without looking at its context.

Now this open proclamation of a 'contextual approach' serves a number of purposes. It acts as signal to lawyers that Lord Steyn is not in the camp of the arid formalists or the naïve literalists. But that camp has surely long since been abandoned, and may never have been occupied at all except by straw men or hermits in ivory towers. It also acts as a signal that he shares the impatience of most modern judges with the intricate games of the parol evidence rule, which excluded 'extrinsic evidence' about the meaning of documents except in numerous special cases, the exceptions eventually becoming so numerous that the Law Commission concluded that the rule no longer existed for all practical purposes.[9] Lord Steyn is also associating himself with late twentieth century philosophies of meaning, which, in reaction to Viennese logical positivism, insisted that words obtained their meaning from their usage in ordinary or specialist language.

[7] Steyn, above n 1, p 125.
[8] Steyn, above n 1, pp 124–125.
[9] Law Commission, *Law of Contract: The Parol Evidence Rule*, Report No 154, Cmnd 9700 (1986).

Yet Lord Steyn does not explain in detail what this contextualism involves. It is often the decisions about which context is relevant, rather than whether context can be relevant, which divides judicial opinion today. The commitment to contextualism is in fact opaque. The literal meaning of context is 'not text'. The prefix 'con' means 'not', a negation of what comes afterwards. Context means literally all things that are not the text. To be a contextualist, therefore, is to say that in interpreting texts, one should examine everything, or at least some things, which are not the text. For the purposes of interpretation, we are told, text is what matters, but apparently the text can only be rendered meaningful by ascertaining what is not-text or context. How much, therefore, does text matter, if we can only know what it means when we know what it is not?

In the remainder of his chapter, Lord Steyn applies this contextual approach to the fields of interpretation of contracts, statutes, and constitutional documents including bills of rights. Without rehearsing all his interesting remarks here, I shall merely draw out a few points that are connected to the question of how much does the text matter to a committed contextualist.

In connection with the interpretation of written contracts, it is worth examining further Lord Steyn's observation that the context may be significant in suggesting that a fairly literal approach to a text should be adopted. When examining a documentary credit, for instance, the court perhaps should look at nothing beyond the four corners of the text, that is the literal or obvious meaning.[10] Here, context not only determines the meaning of the text, but also determines the degree of latitude with the text in the attribution of meaning, including, in some instances, a direction not to look beyond the document itself, that is to ignore context. As an additional point about the interpretation of contracts, Lord Steyn also gives a poor prognosis for the two 'sacred cows' of contract interpretation that exclude the admissibility of evidence about negotiations and post-contract performance.[11] This kind of evidence appears to be an appropriate and relevant context to consider. He shows how, in any event, the courts avoid these limitations in practice by various devices that restore the relevance of context, such as estoppel by convention.[12]

Despite being a contextualist, in his discussion of statutory interpretation, Lord Steyn subjects the decision in *Pepper v Hart*,[13] the authority that permits reference to Parliamentary debates reported in *Hansard* in order to ascertain the meaning of a statute, to that familiar legal technique of confining it to its own special facts.[14] He insists that ministerial statements

[10] Steyn, above n 1, pp 125–126.
[11] Steyn, above n 1, pp 128–129.
[12] *Amalgamated Investment & Property Co Ltd v Texas Commerce International Bank Ltd* [1981] 3 All ER 577, CA.
[13] [1993] AC 593, HL.
[14] Steyn above n 1, pp 131–136; cf J Steyn, '*Pepper v Hart:* A Re-examination' (2001) 21 *Oxford Journal of Legal Studies* 59.

should not usually be considered, partly because that strengthens the position of the executive over Parliament, and partly because that practice ignores the principle that 'text matters'. This view seems odd coming from an ardent contextualist. After all, given the reality of the executive's domination of Parliament, a ministerial statement about the purpose of a statutory provision is likely to be the most reliable kind of context available. This criticism of *Pepper v Hart* seems to be another example where the context tells the judge not to look at the context.

On the other hand, with respect to constitutional adjudication, Lord Steyn argues that courts should not be afraid of giving constitutional rights a strong and broad interpretation with a view to strengthening democracy. Here we might infer that the context of bills of rights suggests that the text does not matter so much, and that judges should adopt an expansive interpretation of rights.

My comments about Lord Steyn's discussion of contextualism reveal, I think, the tension between his two central propositions about interpretation of documents. 'Text matters', to be sure, but how much it matters, depends on context. Thorough-going contextualism is driven to the position that context determines how much text matters. What is not-text determines how much weight should be attributed to the text. 'Text matters' is not a basic principle of interpretation in a contextualist approach; it is a product of the context. The context may determine that the text is all-important, or of little importance at all, or somewhere in the middle. For the committed contextualist, ultimately it is only context that really matters.

II. THE DISUNITY OF INTERPRETATION

My second point follows from the previous examination of committed contextualism. A committed contextualist cannot believe that a unified approach to the interpretation of documents is possible or in fact practised. If text only matters to the extent that the context determines that it should, the context of the type of document will inform us as to how much weight should be placed on the words used in the text. Committed contextualism has to accept that the approach to the interpretation of contracts, statutes, and constitutional rights must differ in important ways because the context differs. Lord Steyn does precisely what a committed contextualist should. Despite purporting to be discussing interpretation in general, he divides his discussion into three parts and argues in favour of different approaches towards legal texts. Indeed, as we have noted, he argues for disunity in the proper approach towards interpretation within categories. In the case of the interpretation of contracts, the text should matter far more in a documentary credit than in a standard form consumer contract. This difference arises perhaps because in the case of documentary credits it is the text itself that is being acquired, not the underlying assets. It is worth exploring this implication of committed contextualism rather further, by considering three contrasting approaches in practice between the interpretation of contracts and of statutes. These contrasts hark back to my original puzzle about Lord Blackburn's attitude towards the relevance of intention in construing legal documents.

A. Intention

It is commonplace for judges to assert that their job in interpretation is to ascertain the 'intention of Parliament' and the 'intention of the parties to a contract'. That is what Lord Blackburn said, and it is surely the most often repeated assertion about interpretation of contracts and statutes.

> In construing this provision, as in any other contractual provision, the object of the court is to give effect to what the contracting parties intended. To ascertain the intention of the parties the court reads the terms of the contract as a whole, giving the words used their natural and ordinary meaning in the context of the agreement, the parties' relationship and all the relevant facts surrounding the transaction so far as known to the parties.[15]

In connection with the interpretation of contracts, Sir Christopher Staughton has described the requirement to look for the intention of the parties as 'Rule One'.[16]

It is also true, of course, that in conjunction with this assertion of the importance of intention, the proposition that 'text matters' is awarded considerable weight. In the case of commercial contracts, the job of the courts is to ascertain the intention of the parties from the language which the parties used in their contract. Similarly, the starting-point for the ascertainment of the intention of Parliament must be a reading of the words of the statute itself. This similarity was the basis for Lord Blackburn's assertion of the unity of approach to interpretation of legal documents. Yet this similarity in approach to the connection between intention and the text seem to me to obscure a significant difference in practice between the interpretation of contracts and of statutes.

For the purposes of statutory interpretation, the courts are engaged in trying to construct the actual intention of Parliament. In general, the courts restrict their search for intention to the words of the statute, because they might discover all kinds of conflicting intentions were present in the Members of Parliament when they voted for a particular piece of legislation.

[15] Lord Bingham, *Bank of Credit and Commerce International SA v Ali* [2001] UKHL 8, [2002] 1 AC 251, para 8. In the same paragraph, however, Lord Bingham also endorses the speech of Lord Hoffmann in *Investors Compensation Scheme Ltd v West Bromwich Building Society* [1998] 1 WLR 896, 912.

[16] C Staughton, 'How Do the Courts Interpret Commercial Contracts?' (1999) 58 *Cambridge Law Journal* 303, 304.

Yet the process of interpretation of statutes is one that is designed to ascertain the purpose of the enactment, which assumes that the objectives of the legislators should provide the ultimate guide to the meaning of the legislation. This reference to purpose, which involves the hypothesis of a collective intention, provides the reason why in *Pepper v Hart* the House of Lords accepted that in some circumstances it must be correct to refer to ministerial statements that explained the objective of the sponsor of the legislation. The purpose or intention of the legislation is recognised as the crucial context for its interpretation. For pragmatic reasons regarding the length and expense of litigation, we may wish to confine detailed trawls through the proceedings of Parliament in order to discover some stray remark that may throw light on an obscure provision. Nevertheless, ultimately the task of the judge is to give effect to Parliament's purpose, which may involve giving the words of the statute either a narrow or an expansive construction.

Intention seems to have a different relevance in connection with the interpretation of contracts. The objective approach insists that actual intention is in principle irrelevant. There are certainly pragmatic reasons for not launching an enquiry into the actual intentions of the parties. For a start, this enquiry is likely to be fruitless, because it will only reveal that the actual intentions of the parties differed. That disagreement is probably why the parties are before the court. Even if their intentions had coincided when the contract was formed, the incentive to insist that there was a different intention is likely to prove too strong a temptation in order to take advantage of the circumstances that have subsequently occurred. But aside from these pragmatic considerations, there is a reason in principle not to examine the actual intentions of the parties when interpreting contracts.

The 'objective approach' to the interpretation of contracts holds that the meaning which should be attached to the words must be guided by what the reasonable promisee would have understood the words to mean. This test does not examine the intention of the promisor, but rather examines the impression that the words formed on the mind of the hypothetical reasonable promisee. In principle, the context of the intention of the promisor is irrelevant to this enquiry. The crucial context is how a reasonable person would have understood the words in the contract. A reasonable person would no doubt take into account the likely purposes of the promisor in forming a view about the meaning of the words used in the promise. A reasonable person would also be influenced by additional evidence that indicated what the actual intention of the promisor had been. In the cases of 'snapping up' of mistaken offers, the courts insist that actual knowledge of the intention of the promisor must influence the reasonable promisee's understanding of the words used in the offer.[17] But in the absence of any further communication about the actual intention of the promisor,

[17] *Hartog v Colin & Shields* [1939] 3 All ER 566, QB.

the objective test insists that it is not the intention of the promisor which is relevant, but the impression that the words would form on the mind of the reasonable promisee.

Let me illustrate this distinction with respect to the relevance of intention by reference to the example used by Lord Steyn, where he draws on Wittgenstein's remarks about meaning in his book *Philosophical Investigations*.[18] His example is where a parent, on going out, tells a carer to 'teach the children a game'. The carer teaches the children poker, and on the parent's return, the parent is outraged and says that is not what the parent meant at all. This example is normally used to point out that literal meanings of words like 'game' are not always the meaning in context; in other words, meaning depends on context, or, as Wittgenstein argued, on use in a 'language game'. In some contexts, poker is a game; in others not. The contrast I am trying to draw, however, is a different one.

In one context or language game the intention of the speaker, in this case the parent, is crucial: poker is simply not a game within that framework. The parent did not mean by the word 'game' various types of things that might be regarded as games in other contexts, including poker. But if we change the framework and ask what the 'reasonable carer' would have understood the word 'game' to mean, what the parent actually intended is irrelevant, and what becomes important is what a reasonable carer might view the word 'game' as meaning. A reasonable carer would no doubt impute some kind of intention to the parent for the purpose of interpreting the instruction, such as keeping the children entertained, but the actual intention of the parent is not relevant to this enquiry. My suggestion is that this difference in the relevance of intention points to a fundamental contrast between statutory interpretation and contractual interpretation.

For the purposes of statutory interpretation, the language game is similar to that involved in the example about the parent giving instruction to the carer. Parliament issues orders in the form of statutes, and the courts are expected to try to understand the meaning of the statutes by reference to the purpose of Parliament in issuing the orders. To ignore purpose in the interpretation of statutes would be to commit the error of the carer, which was to misunderstand the language game involved in the use of imperative language. The carer used the dictionary definition of the word game, which would certainly be relevant in other contexts, but here, where an instruction is being issued, the language game requires close attention to the purpose of the speaker in order to confine or expand the meaning of the words. But is the language game the same in the context of the interpretation of contracts?

[18] L Wittgenstein, *Philosophical Investigations* trans G Anscombe (Oxford: Blackwell, 1958) note to para 70; Steyn, above n 1, p 124.

The objective approach to the interpretation of contracts suggests the contrary. The intention of the promisor is only relevant to the extent that suppositions about intention inform what the reasonable promisee might reasonably have supposed was being promised, the reasonable expectation arising from the text. What the objective test proclaims is that contract interpretation involves a different kind of language game. The text matters more than in the case of imperative language such as statutes, since the promisee is entitled to place a reasonable construction on its meaning, even though that in fact contradicts the intention of the promisor. Reverting to the example of the carer teaching the children a game, the objective test insists that meaning should be understood from the perspective of the reasonable carer. If the reasonable carer would take the view that the word game included anything that would entertain the children, the carer was entitled to regard poker as a game. My argument leads to the conclusion that the objective approach to interpretation of contracts signals that the language game should be regarded as different from that involved in statutory interpretation. The correct approach to the ascertainment of the meaning of contractual promises does not involve reliance upon the actual intentions of the promisor except to the extent that the reasonable promisee might be influenced by suppositions about the likely scope of those intentions.

This philosophical point about meaning appears rather abstract, but it has important practical ramifications. For example, it explains our difficulty with observing the religious ritual surrounding the first sacred cow. Roy Goode comments that: 'Very often the record of negotiations culminating in the contract is the best guide to the intention of the parties, as is their behaviour subsequently'.[19] He suggests that these legal restrictions are 'widely ignored in practice' (though no evidence is cited). Certainly the courts find a route around the restriction when that helps to resolve the issue.[20] Assuming that it is correct that negotiations provide a good insight into intention, at least in some instances such as the cases involving snapping up of offers, the philosophical point explains our difficulty in abandoning the sacred cow. If actual intention is in principle irrelevant under the objective approach, prior negotiations as a revelation of intention must themselves be in principle irrelevant. That is the justification in principle for the rule. Nevertheless, prior negotiations may well inform the reasonable promisee about the meaning to be attributed to the text. The reasonable expectations of the promisee about the meaning of the text must surely be guided by information arising during the negotiations. In short, the rule is correct in principle under the objective approach, but is stated too broadly. Prior negotiations should be relevant if they would have influenced the reasonable promisee in her understanding of the words used in the contractual document.

[19] R Goode, *Commercial Law* 2nd edn (London: Penguin, 1995) 97.
[20] *The Karen Oltmann* [1976] 2 Lloyd's Rep 708.

My philosophical point also throws light on the considerable stir created by Lord Hoffmann's judgment in the *Investors Compensation Scheme* case.[21] A large part of the fuss about this judgment arises from Lord Hoffmann's stance of committed contextualism. He appreciates that once one accepts that context determines meaning, any aspect of the context may be relevant. He is therefore impatient with any artificial restrictions on references to context, though he does reaffirm the exclusion of prior negotiations, at least to some extent, for pragmatic reasons. Lawyers fear, no doubt with good reason, that such an inclusive approach to the 'matrix of fact' will result in a huge waste of time and money.[22] But this ground for complaint against this approach to interpretation misses what is radical and important about the judgment.

The most important feature of this judgment is the absence of reference to the intention of the parties. Lord Hoffmann instead insists upon the objective approach:

> Interpretation is the ascertainment of the meaning which the document would convey to a reasonable person having all the background knowledge which would reasonably have been available to the parties in the situation in which they were at the time of the contract.

In this passage, and his subsequent propositions, Lord Hoffmann not only implicitly rejects the relevance of actual intentions, but also declines to follow the orthodox statement, that is Sir Christopher Staughton's Rule One, which suggests that the task of interpretation is concerned with intentions (as discovered in the text). What is important is what the reasonable promisee would have understood the words to mean. Background information available to the reasonable promisee would influence that interpretation, and that background information would include information about the purposes of the promisor. But the reasonable promisee must devise a reasonable interpretation of the words, in which background information about the possible intention of the promisor should not be any more significant than other sources of information. The reason why Lord Hoffmann's judgment should be regarded as disturbing is simply because, following the logic of the language game of the objective approach to interpretation as I have explained it, he boldy rejects Sir Christopher Staughton's Rule One altogether.

Matters only get worse for the orthodox views about interpretation of contracts as Lord Hoffmann pursues the logic of the objective approach. He pronounces the heresy that text may not matter in some cases. Where the

[21] *Investors Compensation Scheme v West Bromwich Building Society* [1998] 1 WLR 896, HL; E McKendrick, 'The Interpretation of Contracts: Lord Hoffmann's Re-Statement', in this volume at ch 6.
[22] Staughton, above n 16, 307.

reasonable promisee should realise that the text contains a mistake, such as an error in syntax, a misuse of a word, a Malapropism, a reasonable promisee should attribute a meaning to the text which falls outside any possible dictionary meaning. It may be rare that such mistakes are made, but it remains a possibility, and in some cases the reasonable promisee should realise that a mistake has occurred.[23] The identification of the mistake does not depend upon a reference to the actual intention of the promisor, though certainly that may be relevant. The crucial issue is rather whether the reasonable promisee, with all her background information, should have realised that the promisor had not expressed himself correctly and that a different meaning should be attributed to the words. The implication that sometimes the text does not matter at all, that the courts may override the words used by the parties and replace them with different words, earns the special condemnation of Sir Christopher Staughton,[24] because in truth it rejects all his proposed rules of contractual interpretation. Even the rule that permits the court to reject interpretations of the text that lead to unreasonable results,[25] still insists that the court should find another meaning that the text might bear.

The tension which I have exposed between, on the one hand, the orthodox view that the courts should seek the intention of the parties from the text of the agreement and, on the other, the objective approach that denies that intention is directly relevant at all, is certainly not a new phenomenon. It is bound to erupt from time to time, in my view, once one appreciates that there is an underlying disagreement about the nature of the language game involved in the interpretation of contracts. In the United States the tension flares up in a debate between 'objectivists' and 'subjectivists' about whether proof of a common actual intention of the parties is relevant to interpretation. Judge Learned Hand thought proof of a common intention was irrelevant to the task of assigning a meaning to the terms of the contract,[26] but the prevailing view seems to be that judges should give effect to a common intention of the parties, if it can be discovered.[27] In drafting the *Principles of European Contract Law*, the authors faced the difficult task of reconciling the civilian subjectivist approach with the objective stance of the common law. The synthesis accepts that the discovery of a common

[23] Lord Hoffmann had earlier applied this view in *Mannai Investments Co Ltd v Eagle Star Life Assurance Co* [1997] AC 779, HL.
[24] Staughton, above n 16, 309.
[25] Staughton, above n 16, at p 308, based upon, *inter alia*, Lord Reid: 'The fact that a particular construction leads to a very unreasonable result must be a relevant consideration. The more unreasonable the result the more unlikely it is that the parties can have intended it, and if they do intend it the more necessary it is that they should make that intention abundantly clear' (*Wickman Machine Tools Sales Ltd v L Schuler AG* [1974] AC 235, 251).
[26] *Eustis Mining Co v Beer, Sondheimer & Co*, 239 F 976, 984 (SDNY, 1917).
[27] *Berke Moore Co v Phoenix Bridge Co*, 98 N H 261, 98 A 2d 150 (1953); American Law Institute, *Restatement of the Law of Contracts 2d* (St Paul, Minn, 1981) s 201.

intention is the goal, but also recognises that in the absence of a discovery of a common intention the contract is to be interpreted according to the meaning that reasonable persons of the same kind as the parties would give to it in the same circumstances.[28] It must be doubted whether this synthesis either corresponds to any particular legal system or presents a coherent view of interpretation. It tries to unite those civil law systems that still insist upon a will theory of contract by its reference to the common intention of the parties with the common law's objective approach where intention is of dubious relevance. In so doing, it perhaps changes the objective approach, from one that examines the words solely from the perspective of the reasonable promisee, to one that examines the question from the perspective of both parties, a change that seems necessary to achieve some compatibility with the will theory of contract.

Whatever the merits of this discussion of the method of establishing meaning in contracts by the objective approach, my general point is that, however the objective approach is understood, it involves a different method than that required by statutory interpretation. The intention or purpose of a statute is undoubtedly the primary point of reference in order to resolve ambiguity or vagueness. If the statute prohibits the taking of a vehicle into the park, the central question in borderline cases such as riding a bicycle is: does the conduct fall within the intended purpose of the statute? If a contract provides for the taking of all vehicles out of the park, the central question is what the reasonable promisee understood the word vehicles to mean. The actual aim of the promisor who has requested the removal of vehicles is only relevant in so far as its disclosure would have influenced the reasonable promisee's understanding of the meaning of the word vehicle in the contract.

B. Implied Terms

Lord Steyn describes the use of implied terms in contracts as an aspect of interpretation.[29] I agree that at least in some instances the process of implication of terms is closely related to the activity of interpretation. The ubiquitous 'officious bystander' draws out implicit understandings, or reasonable expectations that have not been explicitly promised in the text. The judge fills in the gaps in the contract by reference to its commercial purpose on the ground of business necessity. There does not seem to me to be an equivalent process in statutory interpretation. Or perhaps it is more accurate to say that if judges are minded to add a subsection to a statute, they do not usually do so openly, but claim rather that the subsection is already present in the text. Why is there this difference, if, as is indicated by

[28] O Lando and H Beale (eds), *Principles of European Contract Law Parts I and II* (The Hague: Kluwer Law International, 2000) Art 5.101.
[29] Steyn, above n 1, p 129.

Lord Blackburn, statutory interpretation and contract interpretation are the same activity?

I suggest that this difference reveals in fact another ground for differentiation between statutory interpretation and contract interpretation. As a broad generalisation, statutory provisions have the basic message 'Don't do this', whereas contract provisions have the basic message 'We want to do this'. The problem of interpretation is to decide the meaning of 'this' in both sentences. In the former case of statutory interpretation, the meaning of 'this' must be given a restrictive meaning, for the liberty of the individual provides a background right from which the statute derogates. So the meaning of 'this' in statutes cannot, or, at least, should not, ever stray beyond the four corners of the literal meaning of the words. In the case of contracts, however, the meaning of 'this' is describing a goal to be achieved and indicating the means to be adopted. Since contracts usually require some element of cooperation between the parties to achieve their goal, it is possible to infer from the document additional obligations that are necessary for the achievement of the objective, notwithstanding their lack of explicit mention in the text. These additional obligations are the implied terms.

I am not suggesting that the courts do not add to the literal meaning of the words in a statute. The difference between statutory interpretation and contract interpretation in this respect is a matter of form rather than of substance. To add a new subsection to a statute, a judge must insist that the duty is already contained within the words that have been enacted. In relation to contracts, however, the judge can acknowledge the existence of a gap and fill it by an appropriate implied term.

But this process of implying terms in fact also indicates that it may differ significantly from the practice of interpretation. To justify the implication of a term, the court relies upon the supposed common intentions of the parties or at least a view of the shared commercial objectives of the transaction. This process only seems similar to the interpretation of contracts if one accepts the view expressed in Sir Christopher Staughton's Rule One, that the courts should search for the common intention of the parties. If, on the contrary, one accepts a thorough objectivist approach, the view that I have attributed to Lord Hoffmann, the techniques differ substantially. On this latter view, interpretation tries to establish what a reasonable promisee would have understood by the text, and the use of implied terms is a radically different judicial technique that helps to articulate and reinforce the commercial purposes of both parties to the contract.

C. Temporality

I want to highlight one further difference between statutory interpretation and contract interpretation. Lord Steyn observes that statutes once enacted have to be interpreted perhaps many generations later, and that judges ought to apply interpretations that fit into the context of the legal system at the time of interpretation, not at the time of enactment.[30] This view is similar to Ronald Dworkin's application of his ideal of 'integrity' in connection with statutory interpretation.[31] It follows that statutes can change their meaning over time. As times change and the legal system with it, the context changes, and so may the meaning of the statute. The word vehicle can at the time of enactment refer to a horse-drawn carriage, but today it includes a motorcar. To reach that conclusion, a court must establish the purpose of the statute at a high level of abstraction, such as protection of public safety in the park.

The same potential to shift meaning with time is less clearly applicable to contracts. It is usually stated that the meaning of the document is to be ascertained by reference to the circumstances at the time of the formation of the contract, that is when the parties committed their agreement to paper. That emphasis upon the moment of formation is why the subsequent conduct of the parties is, in principle, irrelevant, and why that cow is also sacred. It may be necessary to qualify that view about contract interpretation in a long-term relationship where the parties create adaptive possibilities in the contract by conferring discretion on one or both parties. In such cases, such as a contract of employment, the pattern of how the discretion has been exercised may create reasonable expectations about future conduct, which deserve to be protected through implied terms. But that is not really a matter of interpretation; it is more an aspect of good faith in performance, a different use of implied terms.

In sum, statutes have a temporal, contingent meaning, whereas contracts have, at least in principle, a static meaning. Contracts are a form of self-regulation, which can be changed easily by consent, whereas statutes have to be interpreted in ways that fit into the changing times.

D. Documentary Disunity

The three differences which I have drawn between processes of statutory interpretation and contract interpretation support the argument that, in spite of the superficial similarity of the process of attributing a meaning to a text, these two processes have profound differences. My explanation is that legislators and parties to contracts are playing different 'language games' in the sense intended by Wittgenstein. Both games are rather specialized, which takes them outside the usages of ordinary language. Legislation

[30] Steyn, above n 1, pp 130–131.
[31] R Dworkin, *Law's Empire* (London: Fontana, 1986) 348.

fits into a practice of issuing orders to others, in which the intention of the speaker, in this case Parliament, should be regarded as determinative of meaning. Making contracts involves a different practice, one of committing to paper the memory of an agreement. Whether or not agreement was ever in fact reached, whether or not the parties had a common intention, the document is binding under the objective test. The task of the court is to attribute meaning on the basis of what each promisee reasonably expected from the words used to record the transaction.

This difference between the interpretation of statutes and contracts should not surprise a committed contextualist. A contextualist must accept that the nature of the document under consideration must determine how one should approach its interpretation. To think that there should be a common approach to all legal documents would be to reject the essential insight of committed contextualism, namely that context determines how much the text matters.

III. PURPOSIVE INTERPRETATION

In defending the claim that there is a disunity of techniques of interpretation of legal documents, I need to consider a seductive objection that has gained considerable currency. The idea of 'purposive interpretation' has been greeted as the antidote to formalism or literalism. Purposive interpretation requires courts to examine the meaning of a legal document in the light of its purpose. Lord Steyn describes purposive interpretation as desirable for all aspects of legal interpretation.[32] I certainly do not wish to discourage purposive interpretation, but if it is generally applicable to all legal documents, the generality of the method poses in itself a challenge to the claim that there is a disunity of interpretative techniques in law. This challenge is also not merely theoretical.

I have acknowledged that in searching for the meaning of legislation, the courts use purpose as a determining guide. For the task of interpreting contracts, I have conceded at least that for the introduction of implied terms, the courts refer to purpose or commercial object. To that concession, I should add that when the courts reject a possible interpretation as unreasonable, this conclusion may sometimes result from the perception that the proposed interpretation would serve no useful purpose for either party. But I would not like these concessions to be taken to amount to an acceptance that purposive interpretation provides a unified technique for the interpretation of all legal documents.

[32] Steyn, above n 1, pp 125–126.

A. Purpose in Statutory Interpretation

In my view, the use of the word purpose in the context of interpretation of statutes serves as disguised reference to intention. This disguise seems to serve two purposes that are important to a court. First, the term 'purpose' objectifies or, more precisely, 'reifies' intent. Purpose is a thing, which, following excavation, can be examined. And secondly, purpose is always present, waiting to be discovered, even if it can only be stated at a level of abstraction. In contrast, intention may not be there at all, or uncertain, and in any case intention relates to mind not to body, so it can only be the subject of speculation, never examination. The advantage of looking for the purpose of a statute rather than the intention of Parliament is plain.

Parliament always has a purpose, though its intention may be ambiguous, vague, uncertain, non-existent, or controversial. The purpose can be constructed out of a body of material which includes the text, but extends to all the context as well. The judge has the power to determine purpose, and can use this process of discovering purpose in order to attribute a particular meaning to the statute.

Yet the justification for searching for the purpose of a statute seems to me to be the same as the justification for examining the intention of Parliament. As well as acknowledging the subordinate role of the judiciary in fixing policy, it recognises that for the task of establishing the meaning of statute the correct language game is to understand the intention of the person or body giving the orders. The difference between the idea of purpose and the idea of intention does not signal a change in the language game and what determines the meaning of a statute. Instead, the resort to purpose helps to overcome some practical difficulties, such as the case where no clear evidence of intention can be discovered, and enables the courts to overcome temporal problems alluded to above. In other words, for the process of statutory interpretation, judges remain faithful to the task of seeking the intention of Parliament even whilst using the language of purpose.

The search for the purpose of a statute seems also to help the courts to evaluate the relative merits of evidence about parliamentary intention. Lord Steyn, in his essay, seems to resort to purposive interpretation in order to justify his exasperation about the influence of *Pepper v Hart*. It certainly must be tedious for the court to have to listen to all the political posturing in *Hansard,* and he is no doubt right that it rarely throws any light on the point in issue. To explain his unwillingness to listen to all this material, Lord Steyn relies, I suggest, on a distinction between intention and purpose. It is permissible to examine White Papers, official reports, and the like, he suggests, because they explain the mischief or the purpose of the legislation.[33] But investigations of *Hansard* for ministerial statements will only produce rather

[33] Steyn, above n 1, p 135.

ill-considered, unrepresentative, and ambiguous clues to the actual intention of Parliament. All this material only goes to show that Parliament does not really have an intention, because of its composition and procedures. Is this a tenable distinction between 'purpose', which can be attributed to Parliament, and 'intention', which cannot?

My doubt about all this use of a contrast between intention and purpose is that I am not persuaded that the distinction amounts to more than a linguistic device to disguise some problems with respect to the discovery of intention. The language game has not altered. The intention of the legislator remains the determinative influence on the meaning of the words used. The appeal to purpose is made when that intention is not readily discoverable. Using the concept of purpose to justify the exclusion of references to *Hansard* seems to me to be an unpersuasive line of argument. Parliamentary discussions can throw light on both intention and purpose. These terms, though having the different properties described above, share the same view of how interpretation should be accomplished. They signify that the interpreter of the document should be searching for its intended meaning.

There may be good practical reasons for stopping advocates from making incessant references to *Hansard*. But there is no reason in principle to exclude such references in the context of statutory interpretation. In particular, *Hansard* is as relevant (and irrelevant) to the ascertainment of purpose as it is to the intention of the legislators.

B. Purpose in Interpretation of Contracts

In connection with the interpretation of contracts, I suggest that references to purpose serve a different function. They are linked to the perception of the reasonable promisee about the meaning of the contract. If a court can identify an object or purpose of the contract, one which should have been obvious to a reasonable person, any interpretation that contradicts or frustrates that purpose cannot be one held by a reasonable promisee.

> [T]he commercial, or business object, of the transaction, objectively ascertained, may be a surrounding fact ... And if it can be shown that one interpretation completely frustrates that object, to the extent of rendering the contract futile, that may be a strong argument for an alternative interpretation, if that can reasonably be found.[34]

Notice in this observation by Lord Wilberforce that he is careful to distance the concept of 'object', which could be a synonym for purpose, from the actual objective or intention of either party. He permits a search for the

[34] Lord Wilberforce, *Prenn v Simmonds* [1971] 3 All ER 237, 240, HL.

objective of both parties, of the contract as a whole, but not an inquiry into the actual intentions of the parties. This method we have already noted being used in connection with the implication of terms on the ground of business necessity.

In the context of the interpretation of contracts, references to purpose are not a signal for an investigation of the intention of parties except in so far as the law assumes that they must have had some kind of rational purpose for entering the contract in the first place. Instead, references to purpose or object are a technique for constructing what a reasonable promisee should have understood the words to mean.

Assuming these contrasts between the use of purpose in statutory interpretation and contract interpretation to be correct, they demonstrate that, despite the superficial similarity of appeals to purpose in both contexts, there is in fact, as predicted by the theory of committed contextualism, an important difference is the use of the concept of purpose in the two contexts.

IV. THE TEXT OF RIGHTS

My final comment is directed to constitutional adjudication. It is clear that Lord Steyn parts company with Lord Blackburn in this context: interpretation of constitutional rights is not determined by a search for intention, whether or not confined by the text. Such documents as Bills of Rights must be understood in the context of the broader constitutional system. They have to be interpreted in ways that strengthen democracy whilst at the same time placing constraints on an overbearing majority. Their interpretation requires no less than the articulation of a moral and political vision for a liberal democracy.

This context for constitutional rights awards judges considerable power, which has alarmed many opponents of Bills of Rights. In rejecting those grounds for opposition, Lord Steyn is of course right that the Human Rights Act 1998 has not caused it to rain sulphur and brimstone,[35] and I accept that the Act has afforded an opportunity to examine critically and constructively a few murkier areas of English law. But this largely positive report has to be qualified, because some of the fears that the courts might undo social legislation seem to me to be justified. The only case involving the Human Rights Act that I teach in contract law seems to me to represent a worrying harbinger of what the future may reveal.

In *Wilson v First County Trust Ltd (No 2)*,[36] the appellant had entered a six months loan agreement for about £5,000, giving as security her BMW car. When the debtor, Mrs Wilson, failed to pay off the loan on the due date, the creditors proposed to satisfy the debt of nearly £7,500 by selling the car.

[35] Steyn, above n 1, p 137.
[36] [2001] EWCA Civ 633; [2002] QB 74, CA.

Mrs Wilson obtained an interim injunction against sale, and persuaded the judge to reopen the credit bargain on the ground that the APR interest rate of 94.98 per cent was 'grossly exorbitant' and therefore 'extortionate', contrary to the Consumer Credit Act 1974 ss 137–8. The judge reduced the interest rate by about one half, and on repayment of about £6,900 Mrs Wilson retrieved her car. But there remained one other outstanding issue. On appeal, the Court of Appeal decided that there was a small error in the documents in the statement of the 'amount of credit'.[37] Under the Consumer Credit Act 1974 s 127(3), the effect of that error was that a court could not enforce the contract against the debtor or enable the creditor to seize the security. If that conclusion was correct, Mrs Wilson was entitled to repayment of £6,900 with interest (at only 8 per cent). But the Court thereupon stayed the proceedings to hear argument about whether s 127(3) was contrary to the Human Rights Act. Their judgment at the rehearing was, in brief, that s 127(3) interfered with the property rights of the moneylender, that is the security over the car, and that it prevented, contrary to Article 6 of the European Convention, a fair and public hearing by an independent tribunal of the moneylender's civil rights. Such a restriction could only be justified if it was proportionate. The court concluded that it was not. A proportionate restriction would have given the court a discretion to do what was just in the circumstances. Thus s 127(3) was incompatible with the Convention rights and a suitable declaration was made. The upshot was that the moneylender won and was entitled to retain £6,900 plus costs.

My students are surprised that the first human rights case they encounter upholds the right of moneylenders to get their pound of flesh. But I am not surprised. The North American experience reveals that the rich and powerful can use litigation about constitutional rights just as effectively as they use ordinary litigation in order to try to persuade courts to protect their interests. These groups may not always be successful, but that will not discourage them from trying. Where I may lead students astray is in suggesting some critical points about this decision. The statutory scheme in the Consumer Credit Act is elegant and comprehensive. The rules about unenforceability of improperly executed and incomplete documents are clear. It is true, as the Court of Appeal indicated, that *Hansard* does not reveal a ministerial intention about the details of these provisions, though there was some discussion of what agreements should be enforceable, despite defects, if justice so required.[38] So the Court of Appeal said there is no clear answer to its rhetorical question of why did Parliament deny courts the power to

[37] *Wilson v First County Trust Ltd* [2001] QB 407.
[38] This reasoning in which the absence of an explicit ministerial justification for a provision is used to support an inference that perhaps the provision was a mistake or not really intended is surely the kind of reasoning that would alarm Lord Steyn and confirm his worries about the misuse of the discretion afforded by *Pepper v Hart* and was expressly criticised in *Wilson v Secretary of State for Trade and Industry* [2003] UKHL 40, at paras 51–67, 110–118, 139–145.

do what is just? The Court accepted that the purpose of controls over enforcement should be to protect the consumer against informal and inaccurate credit agreements. But the Court argued that this aim cannot justify the inflexible prohibition on the enforcement of some agreements that contain what Parliament regarded as crucial defects.

The main justification for inflexible prohibitions is, of course, that Parliament believes that there is a risk that courts will exercise their discretion badly. This reason for excluding discretion did not seem to occur to the Court of Appeal. The discretion to reopen exorbitant credit bargains is a case in point. Although Mrs Wilson was successful in challenging a 94.98 per cent interest rate on a secured loan, such interventions by the courts seem to be rare.[39] So too, it is intelligible for Parliament to deny a court a discretion in the remedy for defects in the documents, because courts may exercise their discretion in ways that contravene the intention or purpose of Parliament.

But the Human Rights Act gives the court a new weapon. The Court of Appeal used the Act to award itself the discretion over remedies denied to it by Parliament, and proceeded in *Wilson* to grant the pawnbroker everything it asked for.

Fortunately, the House of Lords has now reversed this decision, following an appeal launched by the Secretary of State[40]. The House of Lords rested its decision on the ground that the Human Rights Act did not have a retrospective effect on the application of these provisions of the Consumer Credit Act to transactions entered into prior to 2000. In what were strictly speaking obiter remarks, the House unanimously expressed similar misgivings about the Court of Appeal's decision. It doubted that Article 6, the right to a fair trial, was engaged in this case, though it accepted that the ban contained in s 127(3) on enforcement of security and contractual rights probably engaged Article 1 of the First Protocol concerning peaceful enjoyment of possessions. Nevertheless, the House unanimously agreed that the statutory ban on enforcement of defective credit agreements was justifiable because it satisfied the test of proportionality, and therefore s 127(3) was compatible with Convention rights. Lord Nicholls, giving the leading speech, insisted that in relation to social legislation of this type that the courts should limit their reviewing powers, and should recognise that Parliament has the primary responsibility for deciding whether the means chosen to deal with a social problem are both necessary and appropriate. 'The more the legislation concerns matters of broad social policy, the less ready will be a court to intervene'[41]. In the final outcome, therefore, the social legislation was upheld, and Mrs Wilson kept her car and the money.

[39] I Ramsay, *Consumer Protection: Text and Materials* (London: Weidenfeld, 1989) 348.
[40] Above n 38.
[41] Above n 38, at para 70.

A committed contextualist needs to recognise that in the application of texts about human rights to social legislation that this context may mean that the text matters rather more than usual in connection with statutes. The text matters not merely in confining possible meanings that might be attributed to the legislation, but also in raising the threshold for interventions on the basis of appeals to constitutional rights.

V. CONCLUSION

Committed contextualism, as an approach to the interpretation of texts, does not merely claim that the meaning of words can only be understood by situating them in their context, but more fundamentally suggests that the context reveals the 'language game' in which meaning is established. This insight explains the disunity and heterogeneity of legal interpretation, because different kinds of texts rely for attribution of meaning upon different kinds of language games.

In relation to contracts, the language game awards priority to the meaning of the words as they were, or should have been, understood by the reasonable promisee, given the promisee's knowledge of the circumstances. In relation to statutes, the language game awards priority to the intention of the legislator, which often has to be imputed to the legislator under the rubric of 'purpose'. In relation to constitutional rights, the language game awards priority to the judge in imposing a meaning which presents a morally attractive vision of a liberal democracy. Whether or not a constitutional right should be permitted to override a piece of social legislation involves a formidable task of interpretation for the judge: a purpose has to be attributed to the social legislation, and then under the test of proportionality the judge has to determine whether that purpose is sufficiently compelling to require in this particular case an exception to the moral scheme embodied in the constitutional rights. These three tasks for judicial interpretation have little in common except their difficulty and, of course, the label interpretation.

Part 3

Adapting Commercial Law to Modern Conditions

9

Documents and Contractual Congruence in International Trade

MICHAEL BRIDGE

I. INTRODUCTION

IN AN AGE when developments are proceeding rapidly in the field of electronic shipping documents, it may seem odd to be preoccupied with the role played by paper documents in international trade transactions. My concerns may seem not unlike those of a fifteenth century scholar lamenting the dire condition of modern incunabula at a time when Caxton is turning his printing presses in York. Nevertheless, in the words of Kierkegaard, we may live life by looking forwards but we can understand it only by looking backwards. The problems with which I am concerned are in any case problems that would equally arise in the case of electronic documents. With relatively few exceptions, e-law is virtually the same as old law.

Reasons of convenience and dispatch encourage dealings in documents rather than goods in international trade. Furthermore, many such transactions perform a market function that is distinct from any function associated with the physical use of the goods to which the transaction relates.[1] For international trade transactions to be implemented smoothly, it is important that the documentary character of a transaction such as a CIF contract be integrated with its physical aspect. But integrality is not just an aspect internal to a particular contract. One of the most striking features of international trade is the network of related contracts required to move goods great distances from an originating seller to an end buyer. For trade to work effectively, these individual contracts have to be integrated into a coherent whole. Documentary integrality is therefore both internal to a single transaction, in the way for example that the documentary and physical aspects of a CIF contract must be compatible. It is also germane to a network of contracts, so that CIF contracts and letter of credit and carriage

[1] See Michael Bridge, *The International Sale of Goods: Law and Practice* (Oxford, 1999), pp 26–35.

contracts, for example, must work harmoniously together to sustain the commercial enterprise underpinning them.

I turn now to the internal integrality of the CIF contract, whose features may even now, a century and a half into its recorded history, not be fully understood. In particular, certain features of that contract are difficult to reconcile with contract doctrines, such as present and anticipatory repudiation. For network integrality, I shall start by looking specifically at the contract that arises between an issuing bank (or a confirming bank) and the beneficiary of a documentary letter of credit. My main concern here is to see to what extent a letter of credit is as good as cash—a characteristic that encourages commitment on the part of contracting parties—with particular reference to the question whether a bank is bound to pay against forged documents and documentary nullities. When that is done, I shall consider the comparative treatment of documents under CIF and letter of credit contracts, with a view to seeing how the two contracts are integrated in practice.

II. THE CIF BUYER'S DUTY TO ACCEPT CONFORMING DOCUMENTS

There is a reason why the title of this section is framed in terms of the buyer's duty to accept conforming documents, and not of the seller's duty to tender such documents. My purpose is not to conduct a census of the various documents that a seller must tender or to investigate long-standing problems relating to the types of certain documents the subject of tender, for example, insurance certificate or policy, on board or received for shipment bill of lading. Rather, the focus will be on the difficult question of whether the CIF buyer is bound to accept documents that *appear* to be conforming even if in fact those documents contain misstatements. The argument advanced in this paper is that such documents are non-conforming and that the buyer is not bound to accept them. This question issues out of the more general question whether a CIF seller commits a documentary breach when tendering documents containing misstatements, an offshoot of which is whether such breach (if any) occurs at the point of tender or at some other time. These questions are rooted in the characteristic duality of a CIF contract represented by separate physical and documentary duties bearing on the seller and by separate remedies of rejection and damages arising out of a breach of those duties. Before these questions are developed, a brief description of the nature of a CIF contract is in order.

International sale contracts may contain delivery terms ranging at intervals from delivery at the seller's establishment to delivery at the buyer's establishment, the seller's delivery obligation becoming increasingly onerous the closer delivery comes to the buyer's establishment. It is not uncommon to represent these various possibilities in graphic form with smoking chimneys, dockside cranes, ships and an undulating wine-dark sea. The point about a CIF contract is that it cannot be represented on such a diagram. This is because it is a documentary sale which, according to the settled view, is a sale of goods performed through the medium of documents as opposed to a sale of the documents themselves.[2] The seller has no physical duty to deliver the goods to the buyer, but rather is bound to transfer and deliver documents to the buyer that represent the goods. In the case of two of these documents, which classically are the bill of lading and the marine insurance policy, the aim on transferring the document is to have the buyer succeed to the seller's rights against the carrier or insurer as the case may be.[3] Naturally, most of the documents, and especially the bill of lading, cannot be transferred or delivered to the buyer unless and until the seller performs a physical act, or adopts another's physical act, namely the shipment of the goods.

If one could take an extreme documentary view of the character of a CIF transaction, shipment would not as such be the performance of a physical duty but rather the execution of a precondition necessary for the procurement of the required documents. That said, case law to the effect that a late shipment can be the source of both a physical breach and a documentary breach is inconsistent with this view.[4] A buyer, discovering a late shipment after accepting the shipping documents and so losing the right to reject those documents, may still reject the goods themselves if shipped out of time. Less decisively, another case inconsistent with the extreme view concerns the refusal of leave to serve a writ outside the jurisdiction when the seller committed a breach in the jurisdiction by failing to tender documents in England. The court, exercising a discretion in the matter, was of the view that the real breach was the seller's failure to ship in Sweden that rendered it impossible to procure documents for tender.[5] The extreme documentary view of CIF contracts must therefore be rejected.

[2] *Arnhold Karberg & Co v Blythe, Green, Jourdain & Co* [1916] 1 KB 495, 510, 514. The seller's delivery obligation is not to deliver on shore at the discharge port (see *Comptoir d'Achat et du Boerenbond Belge S/A v Luis de Ridder* (*The Julia*) [1949] AC 293; *Congimex Cia Geral SARL v Tradax Export SA* [1983] 1 Lloyd's Rep 250) nor is delivery effected at the loading port when the goods are handed over to the carrier, since the seller is acting on its own behalf and not as agent for the buyer (*Houlder Bros & Co Ltd v Commissioner of Public Works* [1908] AC 276, 290). The various documents tendered by the seller are moreover substitutes for the goods themselves: see *Tregelles v Sewell* (1862) 7 H & N 574, 582 (insurance policy); *Biddell Bros v E Clemens Horst Co*[1911] 1 KB 934, 956 (Kennedy LJ) (bill of lading). See generally Michael Bridge, *The International Sale of Goods: Law and Practice* (Oxford, 1999), ch 5.

[3] In the case of a bill of lading, the transfer is part of a type of statutory novation, so that the buyer takes on duties under the contract of carriage as well as rights (Carriage of Goods by Sea Act 1992, ss 2–3).

[4] *James Finlay & Co v NV Kwik Hoo Tong Handel Maatschappij* [1929] 1 KB 400 (unless, which seems rather strained, one could recharacterise the seller's physical breach as the culpable failure to enter into a proper contract of carriage with the carrier).

[5] *Johnson v Taylor Bros & Co Ltd* [1920] AC 144.

What then are the buyer's documentary needs? First of all, to take the examples of insurance policy and bill of lading, the buyer needs documents that are effective in the sense that their transfer by the seller gives the buyer direct contractual rights against the insurer and carrier without any further involvement from the seller. Since a wider range of documents than the on-board bill of lading now give the buyer direct rights against the carrier,[6] the buyer's need for direct recourse against the carrier in the event of a breach of the contract of carriage can be satisfied by documents other than the on-board bill. Marine insurance is a more difficult case since the practicalities of modern insurance and the receptiveness of standard form contracts to documents other than the policy itself encourages references in legal texts to the insurance document rather than the policy properly so-called.[7]

Besides the buyer's need for documents that are effective as the vehicles for transferring rights, the buyer needs to know that the correct sorts of rights are being transferred. Given the distance of seller from buyer, the impracticability of contacting carrier and insurer in the day-to-day world of transactional activity and the employment of inexpert banks in the handling of documents, this means that the buyer or buyer's agent ought to be able to tell readily from the face of the document whether the correct rights are being transferred.[8] In addition, the CIF buyer must be able to repose confidence in the veracity of the documents being tendered. There are two ways in which bills of lading may for temporal reasons be non-conforming. They may evidence on their face a shipment out of time. As such, they are perfectly accurate documents but, since they are not the documents called for by the contract, the buyer is not bound to accept them.[9]

A more controversial case of non-conformity concerns the bill that on its face conforms to the contract, in reciting a false shipment date, but that has been falsified in such a way that, had it been accurately dated to show untimely shipment, the buyer could have rejected it for non-compliance with the contract. A complicating feature in the discussion of this second case is that sellers are not always the original shippers causing the goods to be loaded on board and having the status of the first holder of the bill of lading.

[6] See the Carriage of Goods by Sea Act 1992, s 2.

[7] The assignment of a marine insurance policy requires the assignment of the beneficial interest in the policy itself, which is accomplished either by an indorsement of the policy or in any other customary manner: Marine Insurance Act 1906, s 50(2), (3). Assignment of the policy has been described as 'a blend of assignment and novation' designed specially for CIF contracts: Howard Bennett, *The Law of Marine Insurance* (Oxford, 1996), pp 334–35. As a result of changes in the levying of stamp duty, which included the repeal of certain provisions of the 1906 Act, for a document to be a policy it need only carry the name of the assured (or an agent), be signed on behalf of the master and designate the subject matter with reasonable certainty: *ibid*, p 20. A number of derivative documents such as an insurance certificate may therefore qualify as a marine policy for the purposes of the Act including s 50.

[8] This is borne out by *Soon Hua Geng Co Ltd v Glencore Grain Co Ltd* [1996] 1 Lloyd's Rep 398.

[9] *Re General Trading Co Ltd and Van Stonk's Commissiehandel* (1910) 16 Com Cas 95.

It is common in international trade on CIF terms for a seller to be an intermediate seller in a chain with no engagement at all in the process that ends with the loading of the goods on board the vessel.

Because the CIF sale is a documentary sale, the buyer is bound to pay against conforming documents and may not insist on first examining the goods for their compliance with the contract.[10] Apart from its obligations in respect of the documents, the seller also owes certain physical obligations to the buyer, even though delivery of the goods to the buyer is not one of them. The seller must ship the goods on the agreed date or within the agreed period, or adopt some other seller's timely shipment, and these goods must comply with the usual implied obligations of quantity, description, satisfactory quality and fitness for purpose.[11] These physical duties of the seller are however held in suspense until the documentary obligations of the parties to transfer and accept the transfer of conforming documents have been carried out.[12] By paying against conforming documents, the buyer runs the commercial risk of having to pursue the seller to recover his money if subsequently terminating the contract because of a discharging breach of the seller's physical duties. This is a matter of no small inconvenience if the seller is of doubtful solvency or resident in another jurisdiction thousands of miles away. The seller, meanwhile, has the assurance of getting the money in and of dealing with a buyer who will not readily or easily reject the goods and throw them back on him in a distant place where they might only command a distress price.[13] The allocation of commercial risk in this way encourages contractual commitment. Since the buyer has no lien on rejected goods for the recovery of the price,[14] these daunting litigation difficulties may persuade him to accept the goods despite their non-conforming character.

The key to the documentary character of a CIF contract resides in the buyer's duty to accept conforming documents and not to plead present and future physical breaches of the seller that are suspended until the documentary exchange has been completed. If the buyer were permitted to plead a physical breach as a bar to acceptance of conforming documents, this would be inconsistent with the allocation of commercial risk under the contract and would not sit well with the denial of any entitlement of the buyer to examine the goods before having to accept the documents. The CIF contract may not be a sale of documents but it is a contract that is performed through the medium of documents. This aspect of CIF contracts is dealt

[10] *Biddell Bros v E Clemens Horst Co Ltd* [1911] 1 KB 934 (Kennedy LJ).
[11] Sale of Goods Act 1979, ss 13–15, 30.
[12] See discussion below.
[13] Scrutton LJ once warned of the dangers of rejection being used as a lever to extort a price reduction: *Szymanowski and Co v Beck and Co* [1923] 1 KB 457, 467.
[14] *Kwei Tek Chao v British Traders & Shippers Ltd* [1954] 2 QB 459.

with in the leading case, which is the rather unsatisfactory decision of the House of Lords in *Gill & Duffus SA v Berger & Co Inc*.[15]

This case concerned a CIF Le Havre contract for the sale of a quantity of Argentinian bolita beans. The buyer refused to pay against the seller's documentary tender unless it included a post-discharge inspection certificate. The House of Lords had little difficulty in concluding that the buyer was not entitled to resist payment against a documentary tender lacking this document. The seller had therefore tendered to the buyer an adequate set of documents, so that the buyer's failure to pay on documentary presentation entitled the seller to terminate the contract. Lord Diplock, with whose speech all the other members of the court concurred, also went on to assert that it was 'beyond controversy' that a fundamental breach occurred whenever the buyer refused to pay against a presentation of documents 'which *on their face* conform to those called for by the contract'.[16] This statement of a duty to pay against *apparently* compliant documents was unreasoned;[17] it was also unnecessary, since on the facts the seller had not presented documents that contained a misstatement.

Before taking the matter of apparent compliance any further, two further matters flowing from the buyer's refusal to pay should be noted. First, there is the question whether the buyer might lawfully have resisted payment if, in the event, it would have been able to reject the goods themselves and terminate the contract. Secondly, if the buyer were not so entitled, would the existence of any breach of contract claim regarding the goods themselves affect the seller's recovery of damages following on from its termination of the contract in the face of the buyer's fundamental breach?

On the first question, the House of Lords ringingly endorsed the traditional view that a buyer is not entitled at the point of documentary tender to anticipate a future right to reject the goods themselves. This view is sometimes expressed in terms of the right to reject the goods being exercisable once the goods are landed[18] though, more precisely, one might say that it arises once the documentary exchange has been completed[19] since from then on there is no longer any commercial need to inhibit the buyer in the exercise of its right to reject the goods. A buyer knowing after exchange that goods have been shipped out of time and intent on rejecting them for that reason has no cause to examine the goods on shore and therefore no need to await their arrival and discharge.[20] Lord Diplock noted carefully, however,

[15] [1984] AC 382. See Treitel, 'Rights of rejection under cif sales' [1984] *Lloyd's MCLQ 565*.
[16] Emphasis added. Lord Diplock makes two other references to the documents having to conform 'on their face': [1984] AC 382, 391.
[17] Lord Diplock used similar language in the letter of credit case of *United City Merchants (Investments) Ltd v Royal Bank of Canada* [1983] 1 AC 168, 184.
[18] *Kwei Tek Chao v British Traders and Shippers Ltd* [1954] 2 QB 459.
[19] *Gill & Duffus SA v Berger & Co Inc* [1984] AC 382, 395.
[20] This has implications for the tolling of the acceptance period in Sale of Goods Act 1979, s 35: see *Bergerco USA v Vegoil Ltd* [1984] 1 Lloyd's Rep 440.

that on the facts no right to reject the goods would have arisen. In pressing the traditional view, he sharply criticised certain judgments in a High Court of Australia case[21] which he took to support the proposition that a buyer was entitled to reject conforming documents if it should '*subsequently*'[22] be found that the goods themselves did not conform. In this matter, the decision of the House of Lords accords with the orthodox view of CIF contracts that due documentary performance puts the commercial risk on a buyer rejecting goods of having to pursue the seller to recover the price already paid. The decision nevertheless has been criticised by Professor Goode on two grounds.[23] The first is that it is commercially unrealistic to expect the buyer to have to pay the price and then take immediate steps to recover it. The response to this criticism is that very many buyers in such a case will be dissuaded from taking such action against a distant seller when they have no lien over the rejected goods. The allocation of commercial risk in a way that favours the seller will therefore have performed its function of keeping the contract alive, though it will have done so by subverting the exercise of contractual terminations rights rather than by countering them head on.

Professor Goode's second criticism, however, is a stronger one. If the seller is indeed doomed to commit a discharging physical breach respecting the goods once the documents have been taken up by the buyer, why cannot the buyer invoke the doctrine of anticipatory repudiation at the point of tender and decline to pay? It would be a lame response to this argument to assert without reason that the doctrine of anticipatory repudiation has no application to CIF contracts, in the same way as, to take another example, a buyer cannot refuse a notice of appropriation containing details of a ship that could not possibly have loaded the cargo that is being appropriated to the contract.[24] In terms of orthodox contract doctrine, there is no apparent reason to deny the buyer, though it will surely be rare for a buyer to have the evidence ready at the point of a tender of conforming documents and therefore confident enough to accept a repudiation. The buyer, moreover, may not first examine the goods before paying against the documents[25] and may not claim extra time for the evidence to emerge. To that extent, the allocation of commercial risk would remain effective even if the doctrine of anticipatory repudiation were accepted as having a role in CIF contracts.

[21] *Henry Dean & Sons (Sydney) Pty Ltd v O'Day Pty Ltd* (1927) 39 CLR 330 ('one of those submerged cases which lawyers in general have tacitly accepted as being a total loss').
[22] Original emphasis.
[23] *Commercial Law* 2nd edn (Penguin, 1995), 956–57.
[24] *The Vladimir Ilich* [1975] 1 Lloyd's Rep 322. In a sense, the notice of appropriation process operates as a secondary contract to narrow the source of goods within a contract dealing with a broader description of generic goods. If the doctrine of anticipatory repudiation were to apply, there would be the odd appearance of the seller anticipatorily repudiating the main contract of sale by making a narrowing offer, in the form of the notice of appropriation, to enter a secondary contract that is inconsistent with the terms of the main contract.
[25] *Biddell Bros v E Clemens Horst Co* [1911] 1 KB 934 (Kennedy LJ).

The doctrine of *anticipatory* repudiation, nevertheless, is not apt to deal with the cases of late shipment and non-shipment in CIF contracts. Existing law, as seen above, treats shipment under a CIF contract as performance and not merely as a physical precondition to the procurement of the documents that the seller needs in order to perform the contract. If, therefore, shipment is late or does not occur at all, then the breach has already occurred by the time of the documentary tender. Moreover, the CIF seller's obligation of timely shipment is a condition of the contract and, indeed, at times has been treated as part of the description of the goods.[26] To explain away the difficulties created by *Gill & Duffus SA v Berger & Co Inc*, by invoking the old language of dependency of contractual promises and characterising the seller's obligation not as a condition but as an independent covenant,[27] is certainly ingenious but it does nothing for the rationality of the law of CIF contracts to raise a living fossil from the deep.[28] Nor does it adequately explain why the buyer should be entitled to reject the goods for untimely shipment once the documents have been accepted. It would be better to say that parties to a CIF contract impliedly exclude rights of the buyer in respect of the seller's physical performance that would otherwise arise before the completion of the documentary exchange. This exclusion would catch both an existing and an anticipatory repudiation. The presence of an exclusion in implied form may not match the rigorous standards that were expressed for effective exclusion clauses in case law before the passing of the Unfair Contract Terms Act 1977, but, just as that Act for effective purposes is disapplied in the case of CIF contracts,[29] so too the contemporaneous introduction of less rigorous canons of construction and a supposedly more tolerant approach to exclusion clauses should be able to accommodate a well-settled allocation of commercial risk assuming this implied form.

The other question outstanding from Lord Diplock's speech concerns the measure of damages recoverable from a buyer who in breach of contract fails to take up the documents, when that buyer would or could have gone on to reject the goods. At the point of documentary tender, the seller is unable to correct any nonconformity in the goods themselves which, if it does exist, means that the seller at the point of documentary tender is on a collision course with future physical breach. It is often said of a damages award that the aim should be to put a contracting party in the position it occupied before the defendant's breach. The seller in the above circumstances

[26] *Bowes v Shand* (1877) 2 App Cas 455; *Kwei Tek Chao v British Traders and Shippers Ltd* [1954] 2 QB 459, 480–81; *Finagrain SA Geneva v P Kruse Hamburg* [1976] 2 Lloyd's Rep 508, 540–41.

[27] AG Guest (ed), *Benjamin's Sale of Goods* 6th ed (2002), paras 19–159.

[28] For a discussion of dependent and independent covenants and the transformation of the law expressed in cases like *Kingston v Preston* (1773) (as reported in *Jones v Barkley* (1781) 2 Dougl 684) and in Serjeant Williams's notes to *Pordage v Cole* (1669) 1 Wms Saund 319, see Bridge, 'Discharge for Breach of the Contract of Sale of Goods' (1983) 28 *McGill LJ* 867.

[29] Ss 26–27.

is at that point dead in the water. Lord Diplock's concession to the buyer in this position is that the seller's damages 'might, however, fall to be reduced by any sum which the buyers could establish that they would have been entitled to set up in diminution of the contract price by reason of a breach of warranty as to description or quality of the goods represented by the shipping documents ... if those goods had actually been delivered to [the buyers]'. This is the language of an *ex post facto* breach of warranty claim by a buyer who has lost, or chosen not to exercise, the right to reject the goods for a breach of condition.[30] The logic of Lord Diplock's concession, however, goes beyond this case and should also extend to buyers who would have rejected the goods as soon as possible after the documentary tender. In this case, the seller's damages might even be discounted to a nominal sum, which adds up to quite a large incentive to a buyer to act in breach of contract by rejecting conforming documents when confident that it will later be able to reject the goods themselves. At this point, the documentary integrity of a CIF contract is under some stress.

The true nature of the seller's documentary duty erodes the documentary integrity of a CIF contract still further and goes beyond Lord Diplock's concession in so doing. As against Lord Diplock's dictum that the buyer's duty is to pay against documents that on their face conform, there is well-established authority that the buyer is not bound to pay against documents that contain a misstatement. In addition, there are certain other rules applicable to CIF contracts that are difficult, if not impossible, to reconcile with Lord Diplock's dictum.

In the two companion *Tamvaco v Lewis* cases,[31] one concerning a quantity shipped in excess of the amount shown in the documents and the other an amount less, the buyer refusing to pay against a documentary tender was held to have a good defence to the seller's claim since the seller was not ready and willing to perform at the point of tender. The language in these early CIF cases, particularly in the second case, is somewhat imprecise since it refers to the buyer's rejection of the cargo, but the outcomes are consistent with the buyer's right to insist upon documents that are accurate as to the shipped quantity.

The bill of lading in *United Baltic Corporation v Burgett & Newsam*[32] was in a curious form since it recited that the goods were shipped or ready for shipment on a certain date close to the final shipment date, so that the buyer was given no assurance at all that the goods had actually been shipped on time. In a dictum unconnected with the central issue—the acceptability of a bill of lading in this form—Bankes LJ made it clear that a buyer's complaint about a bill of lading that it 'did not accurately represent

[30] Sale of Goods Act 1979, s 53(3).
[31] (1859) 1 El & EL 581, 120 ER 1027 (excess quantity); 1 El & El 592, 120 ER 1032 (short quantity).
[32] (1921) 8 Ll L R 190.

the facts' would be a 'good objection'.[33] The point is made again in the Court of Appeal in *James Finlay & Co Ltd v NV Kwik Hoo Tong Handel Maatschappij*,[34] where Sankey LJ referred to 'one of the fundamental conditions of the [CIF] contract—namely, the condition or obligation to give an accurately dated bill of lading'.[35] To the same effect was the statement of Wright J in the court below that the seller impliedly warrants the accuracy of the date in the bill of lading and the buyer is entitled to reject a bill where shipment is not in fact made in the month recited.[36]

The issue was discussed at greater length in *Proctor & Gamble Philippines Mfg Corpn v Kurt A Becher GmbH & Co*,[37] where the court considered a bill of lading that had been falsely dated in respect of a shipment taking place within the shipment period.[38] It was common ground that the seller was under a contractual obligation to tender a correctly dated bill of lading,[39] but the status of the obligation as a condition or intermediate stipulation was a contentious matter. Kerr LJ was in no doubt that the seller's obligation to tender a correctly dated bill of lading was a condition.[40] He went on to say that the CIF seller was bound by a condition 'that the contents of the documents are true in all material respects'.[41] This is all the more significant in that there was no right in the case to reject the goods for late shipment since the goods had in fact been shipped in time. It is also clear from the case that the intermediate string seller is just as much bound by this condition as the shipper who is privy to the falsification of the date. Nevertheless, since it is the *tender* that is being rejected,[42] the seller will theoretically have an opportunity to make a second and conforming tender, if there is sufficient time left in the shipment period to run the notice of appropriation process again (though the likelihood of this in a standard commodities sales string must be remote).[43] Kerr LJ is clear in saying that the seller's duty is a condition 'in the sense that the buyer is entitled to reject a tender of documents which include an incorrectly dated bill of lading'.

[33] *Ibid*, 191.
[34] [1929] 1 KB 400.
[35] *Ibid*, 416.
[36] (1928) 31 Ll L R 220, 225.
[37] [1988] 2 Lloyd's Rep 21.
[38] More particularly, it had taken place within an agreed extension to the last shipment date. The false dating concealed from the buyer its right to receive a discount against the agreed price.
[39] [1988] 2 Lloyd's Rep 21, 31 (Nicholls LJ).
[40] *Ibid*, 28–29.
[41] *Ibid*, 29.
[42] Acceptance has a much more literal and punctilious meaning in the case of documents than it has in the case of goods.
[43] This is because, under the terms of the contract, the goods will have to have been shipped within the shipment period and a notice to that effect will have to have been given within a stipulated period of shipment and then passed on without delay down a sales string. If documents are rejected it is almost inevitable that there will be no time left to find another cargo coupled with an acceptable notice.

The seller who tenders non-conforming documents to a buyer who rejects them commits a discharging breach only when it runs out of time to make an alternative conforming tender. The discharging breach therefore occurs when the seller's inability to procure conforming documents from an alternative source is crystallised, a matter that is peculiarly difficult to date. The buyer nevertheless has it in its power to deny the seller the opportunity to make a second tender, but it will do so by accepting the documents, hence affirming the contract and losing the right to terminate in respect of a breach of the seller's documentary obligations.

In the court below,[44] Leggatt J appeared to take a position[45] that on its face was closer to that of Lord Diplock: a CIF buyer was bound to pay against documents that on their face were conforming,[46] subject to two exceptions. The first was fraud and the second was the existence of sufficient evidence that the documents were not correctly dated. The second exception was drawn from an express term of the contract[47] that the date of the bill of lading 'shall be accepted as proof of the date of the shipment in the absence of evidence to the contrary'. This express term is somewhat opaque—does it really do anything more than recite the rule that the date on a bill of lading is *prima facie* evidence that the goods were shipped on that date?—and is not to be found in that part of the standard form contract dealing with the tender of documents. Furthermore, it is most oddly expressed in the case as an *exception* to the rule of apparent conformity since it is so destructive of that rule. The evidence exception would make more sense as a *qualification* of the buyer's right to reject documents containing misstatements, in that the buyer is waiving such right where it has no evidence of incorrect dating. In the form in which this express provision is phrased, it bears the additional unattractive features of not stating clearly whether the evidence must be currently available or must merely exist, and of not stating when it must be available or in existence.

As for the fraud exception to the buyer's duty to pay, accepted by both Leggatt J and Lord Diplock, its introduction represents an assumption that documentary tenders under CIF contracts and under letter of credit contracts should be treated alike. In CIF contracts, it would be needed only if the buyer had to pay against apparently conforming documents even if in fact they contained misstatements. This matter will be revisited after letters of credit contracts have been discussed. There will also be a defence to payment on the ground that the documents tendered are nullities. This is implicit in the recoverability as on a failure of consideration of money paid against such documents.[48]

[44] [1988] 1 Lloyd's Rep 88.
[45] It is not clear whether he is merely reciting the seller's main submissions.
[46] [1988] 1 Lloyd's Rep 88, 91.
[47] GAFTA 100, cl 6.
[48] AG Guest (ed), *Benjamin's Sale of Goods* 6th ed (2002), para 19–034.

Further authority protects the buyer beyond the reaches of fraud and nullity and supports the seller's obligation to tender documents that do not contain misstatements. If this obligation is breached, the buyer is entitled to reject the tender. According to this authority, the seller is bound to provide a bill of lading that is 'true and accurate in the material statements that it contains'.[49] In *Henry Dean & Sons Pty Ltd v O'Day Pty Ltd*,[50] the CIF buyer of 'Liverpool wheat-sacks' refused to accept a draft for their price with a tender of documents on the ground that the goods did not fit the contractual description. Apart from certain statements in the case to which Lord Diplock took objection in *United City Merchants*, Knox CJ stated that the buyer was bound to take up only 'proper shipping documents', namely, documents concerning goods 'of the description contained in the contract which have been shipped'.[51] This statement raises the question whether any misstatement in the shipping documents vitiates them or whether the misstatements have to be sufficiently material.[52] Given the strictness of documentary standards in CIF tenders, there may be relatively few misstatements that would not pass a test of materiality. It is submitted that, if documents could be rejected if they told the truth, so too may they be rejected if they conceal by misstatement the truth. Hence, a false statement about the condition of the goods,[53] as well as about the fact or date of shipment, will vitiate the tender. The latter cases fall within the range of the seller's duty to tender a bill of lading that is 'genuine',[54] which is a somewhat inapt word and threatens confusion with the separate notions of forged instruments and documentary nullities.[55] The width of the buyer's defence to an action for non-payment renders it practically unnecessary to consider certain other defences that are absorbed within it, namely, fraud and failure of consideration, since in both cases the documents will almost always contain misstatements.

Support for the seller's duty to tender documents that do not contain misstatements may be sought in certain other rules which are hard to reconcile with Lord Diplock's dictum that the CIF buyer is bound to pay against apparently conforming documents. First of all, there is the rule that permits the award of '*Finlay* damages',[56] where the false dating of a bill of lading

[49] *Hindley & Co Ltd v East Indian Produce Co Ltd* [1973] 2 Lloyd's Rep 515. See also to similar effect *Kwei Tek Chao v British Traders and Shippers Ltd* [1954] 2 QB 459, 480 (Devlin J); *Panchaud Frères v Ets General Grain Co* [1970] 1 Lloyd's Rep 53, 60–61.
[50] (1927) 39 CLR 330.
[51] *Ibid*, 336.
[52] This same word ('material', 'materially') is used in the Uniform Commercial Code to qualify the scope of the fraud exception to the bank's duty to pay under a letter of credit (Art 5–109(1)). See below.
[53] See *Cehave NV v Bremer Handelsgesellschaft mbH* [1976] QB 44.
[54] See *James Finlay & Co Ltd v NV Kwik Hoo Tong Handel Maatschappij* [1929] 1 KB 400, 408 (Scrutton LJ).
[55] See discussion below under letters of credit.
[56] See *James Finlay & Co v NV Kwik Hoo Tong Handel Maatschappij* [1929] 1 KB 400 and related cases; Michael Bridge, *The International Sale of Goods* (OUP, 1999), paras 9.99–9.118.

conceals from the buyer the fact of an untimely shipment so that, by the time the truth emerges, the buyer has disposed of the goods and in so doing has accepted *the goods* for the purpose of section 35 of the Sale of Goods Act. The false dating of the bill of lading *causes* the buyer to act in such a way as to lose the right to reject the goods. This is because the buyer acts in a way that is inconsistent with the seller's ownership or retains the goods beyond a reasonable time (usually the former). The consequent effect is that the buyer is prevented from reversing market decline occurring between the contract and delivery dates, which a terminating buyer is able to do.[57] The amount of this market decline constitutes *Finlay* damages. The rule certainly supports the existence of a duty to tender accurate documents. The status of that duty, as a condition or a mere warranty, is however a different matter, though nothing turns upon the classification of the duty in the context of a *Finlay* award. It is precisely because it is too late for the buyer to terminate the contract at all—since having accepted the documents it goes on to reject the goods—that the buyer is awarded damages to mimic the economic effect of termination. And whether the buyer recovers such damages for breach of a mere warranty, or for breach of a condition treated *ex post facto* as a warranty, is neither here nor there. Moreover, a buyer who is aware of the documentary breach at the time of tender would, even if having to accept the documents because the seller has committed only a breach of warranty as regards the documents, not suffer a *Finlay* loss because it would be in a position to avoid accepting the goods under section 35.

A second rule arises out of the controversial case of *Panchaud Frères v Ets Grain Co*.[58] The rule, so far as it can be coherently expressed, is that a CIF buyer accepting documents may not reject the goods for untimely shipment when those documents reveal the fact of untimely shipment, if in the circumstances the buyer by its conduct is estopped from asserting a claim for, or has elected not to terminate the contract because of, late shipment. The case concerned a late shipment where a bill of lading was dated 31 July but the inspection certificate issued at the loading port referred to samples being drawn 10–12 August. The buyer purported to reject the goods for reasons of description but did not raise late shipment until, more than two years later, it was in front of the arbitral Committee of Appeal. In upholding the Committee of Appeal in its decision that the buyer had lost the right to reject the goods for late shipment, the Court of Appeal did not rest its decision upon the time that elapsed after the completion of the documentary exchange.[59] Nor did it as such repudiate the entitlement of a contracting

[57] This is because the buyer may recover the original price (reflecting the market price at the contract date) if paid or, if it has not been, resist paying it.
[58] [1970] 1 Lloyd's Rep 53.
[59] But the Court of Appeal did uphold the arbitral Committee on an additional ground for its decision, namely, that the buyer had raised the objection of late shipment outside the six month limitation period in the arbitration clause.

party to terminate a contract where incorrect grounds for termination are expressed at the time but correct grounds for termination do exist.[60] Instead, the buyer lost the right to reject the goods because it had the opportunity to read the documents prior to accepting them and, had it done so, would have discovered the fact of late shipment. In the words of Lord Denning, the taking up of the documents prevented the buyers from 'rejecting the goods on the ground of late shipment or defective bill of lading'. At another point, he says that the buyer was 'precluded afterwards from complaining of the late shipment or of the defect in the bill of lading'.[61]

The language of Lord Denning confuses the separate processes of rejecting the bill of lading and rejecting the goods themselves. In particular, a defect in the bill of lading cannot as such be a ground for rejecting the goods. The buyer rejects the goods for late shipment and not for a defect in the bill of lading. If the buyer rejects the bill of lading, the buyer does not have to reject the goods because the CIF seller is never in a position to tender goods directly under what is a documentary sale.[62] The important point in *Panchaud Frères*, however, is that the right to reject the goods was lost when tender of the documents was accepted.[63] Indeed, it would not be inaccurate to record the case as saying that, where the documents tendered give notice to the buyer that the goods were shipped out of time, there are not two rights of rejection (of the goods and of the documents) but only one right, and that right is a right to reject the documents. A strained reading of the case might say that it is not inconsistent with the existence of the two rights, but that the buyer in accepting the documents must do so under protest, reserving its right to reject the goods themselves, and reject them specifically for late shipment. But the report does not in any way positively support such a reading, which anyway is so unusual that it ought to have been expressed if that indeed is what the court meant. Taking therefore the straightforward reading of the case, the buyer seeking to terminate the contract should have rejected the seller's tender of the documents. The tender of a falsely dated bill of lading was therefore the breach of a contractual condition to the extent that the buyer was entitled to reject the documents.[64]

III. THE BANK'S DUTY TO PAY AGAINST CONFORMING DOCUMENTS UNDER A DOCUMENTARY CREDIT

The nature of the duty resting upon a bank that issues a documentary letter of credit, or confirms such a letter, is well-known. It arises under the terms

[60] A rule upheld on many occasions and usually associated with the decision of the House of Lords in *British & Benington's Ltd v North West Cachar Tea Co* [1923] AC 48.
[61] [1970] 1 Lloyd's Rep 53, 58.
[62] See eg, *Orient Co Ltd v Brekke & Howlid* [1913] 1 KB 531.
[63] *Ibid*, 60 (Winn LJ: 'accepting these documents with that notice which they had').
[64] Above 57, 60–61.

of a unilateral contract, the inference of which is made in such a way as to ease commerce and provide the buyer with the assurance it needs to proceed with the performance of the contract. Contract doctrine is not allowed to impede the due workings of the letter of credit system, with the result that the very issuance of the credit—and the same can fairly said for its confirmation—has with remarkably little fuss been held to give rise to a letter of credit contract binding on the bank,[65] difficulties of acceptance and consideration being simply swept away.[66]

What then does the beneficiary of a letter of credit need? First of all, the beneficiary needs a solvent and reliable paymaster. This means that a CIF seller knows that shipment of the goods and the procurement of the documents specified in the letter of credit will guarantee payment. This goes beyond the protection afforded by a seller contracting on cash on delivery terms. A CIF seller having to dispose of goods currently on the high seas or at the port of discharge, because of the buyer's insolvency or repudiatory breach, will have to dispose of the goods, probably at a distress price, and will incur various expenses in the nature of agency and carriage costs that may not readily, or at all, be recovered from the buyer in a damages action. The obligation to pay of a solvent bank, particularly a confirming bank in the seller's own jurisdiction,[67] means that the seller will not incur these additional costs. Indeed, the letter of credit gives further assurance. A seller, once the beneficiary of a binding credit, may incur costs in preparing for performance, which might consist of the expense of starting up a factory assembly line or entering into a binding commitment with its own supplier, which in turn may mean making banking arrangements to provide that supplier with documentary credit protection.

Now, the beneficiary's protection may be less than perfect under the documentary credit system. First of all, as unlikely as this may be, the confirming bank may fail and the distant issuing bank may default, thus creating a litigation headache. Secondly, the doctrine of exact compliance of the documents with the documentary requirements laid down in the letter of credit may work against the seller. It is a commonplace that documentary tenders are very frequently defective,[68] which puts the beneficiary at the mercy of the bank, or the applicant for the credit (the buyer) consulted by the bank,

[65] The expression 'letter of credit contract' is used in this paper to signify either or both of the contracts concluded between the beneficiary of a credit and the issuing and confirming banks.
[66] *Hamzeh Malas & Sons v British Imex Industries Ltd* [1958] 2 QB 127, 129.
[67] The bank that takes up the documents and through which payment is made is known under UCP 500 as the nominated bank. It may or may not be the confirming bank but very often is: see *Banco Santander SA v Bayfern Ltd* (Langley J 9 June 1999).
[68] In an article published in an ICC newsletter in winter 2000, Professor Ronald Mann, having conducted a survey of a cross-section of banks, found that in only 27 per cent of cases was there exact documentary compliance with the letter of credit and in many cases there were 'surprisingly serious' discrepancies. See http://www.iccwbo.org/home/news_archives/2001/dcpro_insight.asp (visited 22 January 2003) for the reasons he suggests for the continuing use of letters of credit.

so that the assured payment obligation becomes in reality a payment discretion. It is surprising that the documentary credit system works at all in view of the great frequency of documentary discrepancy.

Thirdly, although the beneficiary of a credit is protected by the doctrine of autonomy—which means that the bank is not entitled to look through the documentary credit contract to the underlying sales transaction—a major uncodified exception to this doctrine arises in cases where the beneficiary is guilty of fraud. The exception is uncodified in the sense that there is no express provision made for it in the Uniform Customs and Practice for Documentary Credits 1993 (UCP 500), which are incorporated in most documentary credit contracts. This omission is undoubtedly due to the difficulty of determining what fraud means and measuring its scope, a matter of uncertainty that further undermines the assurance that a beneficiary might expect from the issuance and confirmation of a letter of credit. The recognition of a fraud exception and the determination of its incidents are left to the law applicable to the documentary credit contract.

As for the principle of autonomy, this is expressed in Article 4 of UCP 500 in terms of the parties to credit operations dealing only with documents and not with the goods or services to which the documents relate. Article 3 stresses that the letter of credit contract is 'separate' from the underlying contract, the bank in no way being 'concerned with or bound' by the latter contract. This principle is equally accepted in English law. A bank is no better placed to object to the performance that has been rendered under the underlying contract than is the acceptor of a bill of exchange or maker of a promissory note when payment is sought by a holder in due course.[69] Indeed, Rowlatt J once went further in his desire to demonstrate that the letter of credit contract was in no way dependent upon the underlying contract. He remarked: '[S]o far from the letter of credit contract being qualified by the contract of sale, the latter must accommodate itself to the letter of credit'.[70]

In this paper, I cannot cover all aspects of the law relating to letters of credit and, indeed, shall only deal, though in some detail, with the problem of forgery and nullity. Hence, I shall not examine the meaning of fraud, nor shall I discuss the nature of the case that has to be made if a bank is to be enjoined from making payment to the beneficiary of a letter of credit or to another bank that has already made payment. Nor shall I consider the circumstances in which a bank may resist summary judgment on a claim brought under the letter of credit. I do, however, want to pay some attention to the notion that fraud must be the fraud of the beneficiary. It is also important to identify the victim of the fraud—whether it is the applicant for the credit or the bank. The loss to a beneficiary may be the market loss that comes with paying for July goods when in fact August goods were

[69] *Power Curber International Ltd v National Bank of Kuwait SAK* [1981] 1 WLR 1233, 1241.
[70] *Urquart Lindsay & Co v Eastern Bank Ltd* [1922] 1 KB 318, 323.

shipped, or it may be the loss arising on the receipt of goods that bear little relation to goods of the contractual description. The loss incurred by an issuing bank, for example, may be the reduced value of its security if it is taking a pledge of the documents as security for an advance to the buyer applicant.[71] This loss, however, is contingent if the buyer is good for the money in any event. The recognition that there may be more than one victim of the fraud[72] may account in part for the proposition debated at some length in the United States that the fraud may repose in the documents or be facilitated by their presentation.[73]

Before turning to forgery and nullity, some preliminary points should be made. If the victim of the fraud is the applicant, who is bound by an underlying contract to the beneficiary, then it makes sense to speak of fraud as an exception to the autonomy principle. If, for example, the victim is the confirming bank, the bank's defence to payment based upon the fraud of the beneficiary should not be seen as an exception to the autonomy principle at all.[74] The basis upon which the fraud exception is recognised should also be understood. It has been justified on the ground that fraud unravels all,[75] which represents Lord Diplock's translation of the maxim *ex turpi causa non oritur actio*, when *fraus omnia corrumpit* would be a more accurate source of the translation. Blanket statements of this kind, however, envelope better than they explain. A different approach, that it is an implied term of the letter of credit contract that the beneficiary shall not be guilty of fraud,[76] is more precise but inapt to deal with the relationship between applicant and beneficiary, the autonomy of whose underlying contract repels any reference to the conduct of the beneficiary in presenting documents pursuant to a letter of credit.[77]

[71] See Ackner LJ in *United City Merchants (Investments) Ltd v Royal Bank of Canada* [1981] 3 All ER 142, 170: 'A banker … ought not to be under an obligation to accept or pay against documents which he knows to be waste paper. To hold otherwise would be to deprive a banker of that security for his advances, which is a cardinal feature of the process of financing carried out by means of the credit … '.

[72] UCC Art 5–109(a).

[73] *Ibid*. The latter case may represent a contraction of its earlier counterpart in former Art 5–114(2) ('fraud in the transaction'), on which see *Rockwell International Systems Inc v Citibank NA*, 719 F 2d 583 (2nd Cir 1983); *Federal Deposit Insurance Corp v Bank of San Francisco*, 817 F 2d 1395 (9th Cir 1987); *United Bank Ltd v Cambridge Sporting Goods Corp*, 392 NYS 2d 265. For Canada, see *Bank of Nova Scotia v Angelica Whitewear Ltd* (1987) 36 DLR (4th) 161, 176 (Le Dain J): '[T]he fraud exception to the autonomy of letters of credit should not be confined to cases of fraud in the tendered documents but should include fraud in the underlying transaction of such a character as to make the demand for payment under the credit a fraudulent one.'

[74] In rare cases, the bank may be induced by the misrepresentation of the beneficiary to enter into the letter of credit contract and will consequently have a defence to payment: *Safa Ltd v Banque du Caire* (CA 20 July 2000).

[75] *United City Merchants (Investments) Ltd v Royal Bank of Canada* [1983] 1 AC 168.

[76] *Czarnikow-Rionda Sugar Trading Inc v Standard Bank London Ltd* (Rix J 6 May 1999).

[77] Autonomy should work both ways. If the letter of credit contract is autonomous of the underlying contract of sale, so too the underlying contract of sale must be autonomous of the letter of credit contract.

The question to be considered now is whether, going beyond fraud, the tender of invalid documents—in the sense of forged documents or documentary nullities—should absolve the bank from having to make payment. The bank's need for a valuable security gives it an interest in the matter independent of that of the applicant. This question of forgery and nullity is closely related to the definition of fraud but should not be seen as bound up exclusively with fraud: the bank's security is rendered worthless if the documents are nullities, and documents may be nullities, as we shall see, even if the fraud exception to payment does not apply. The introduction of forgery and nullity also permits parallels to be drawn with the position of CIF buyer and seller *inter se* that go beyond parallels that might be drawn by focusing only upon fraud. Now, if documents are tendered and they are either forgeries or nullities or are both, then in many cases there will also have been committed fraud and that fraud may fall within the fraud exception. Nevertheless, despite language used in some of the cases, fraud on the one hand and forgery and nullity on the other hand are analytically separate; the latter are not variations of fraud. As we shall see, both a forgery and a nullity exception to payment by the bank have been rejected by the English courts, though the rejection of the latter is questionable.

Before proceeding further, a clarification of the meaning of forgery and of nullity is required. Drawing upon the Forgery Act 1913,[78] Stephenson LJ in *United City Merchants (Investments) Ltd v Royal Bank of Canada*[79] defined a forged document as one that told a lie about itself. It might be a lie about the time or place of its making or it might be a lie about the person who made it. A forgery would therefore include a falsely dated bill of lading and a bill of lading with a falsified signature. The bill would be a nullity if the signature were falsified but not if the date were falsely stated.[80] If the document, however, told a lie about its contents—as a bill might if it misstated the quantity or description of goods shipped—then it would not be a forgery and (one might add) it would *a fortiori* not be a nullity, subject to the following case. Going beyond the illustrations of Stephenson LJ, there is the case of the bill of lading representing a phantom shipment. This is not just a bill that tells a lie about its (non-existent) contents. It is indeed not a bill of lading at all, regardless of the accuracy of the date or of any signature, since it represents no cargo, is not a document of title and gives the holder constructive possession of nothing. It can only be a nullity, even though it seems not to be a forgery. To conclude on these definitional matters, some forgeries are also nullities but some are not, the latter being the narrower case. Yet one example, that of the phantom shipment, concerns a nullity that is not a forgery.

[78] S 1(2).
[79] [1981] 3 All ER 142, 159.
[80] Relying upon *Gian Singh & Co Ltd v Banque de l'Indochine* [1974] 1 WLR 1234, 1238–39, and *Kwei Tek Chao v British Traders & Shippers Ltd* [1954] 2 QB 459, 476. See also Ackner LJ in *United City Merchants (Investments) Ltd v Royal Bank of Canada* [1981] 3 All ER 142, 170.

It is commonly said of documentary credits that they should be as good as cash. To admit of a fraud or any other payment exception at all is to diminish their cash quality; such diminution will have to be kept within narrow bounds if cash quality is to be anything more than pious aspiration. Bills of exchange have a similar cash quality[81] and, indeed, to say of a letter of credit that it ought to be as good as cash is to say that it must be 'treated like a bill of exchange'.[82] It is interesting to see just how far the beneficiary of a letter of credit is treated like the holder in due course of a bill of exchange. First of all, the rights of a holder in due course of a bill of exchange are not unlimited. In the case of forged signatures on the bill, the forgery is wholly inoperative and no right to retain the bill, give a discharge for it or enforce payment under it may be acquired through that signature.[83] So far as the developing law on letters of credit fails to recognise a forgery exception, it diverges at this point from the law on bills of exchange. In addition, there is the following clear difference between letters of credit and bills of exchange. The holder of a bill of exchange can transfer the bill to a holder in due course in such a way that the latter takes clear of personal defences and defects of title of prior parties.[84] A defect of title includes, for example, fraud and duress[85] and excludes the so-called real defences, which include *non est factum*, forgery and the material alteration of a bill. If the beneficiary of a letter of credit discounts it, as might happen in the case of a deferred payment credit, then the discounting bank will be treated as an assignee and, like any other assignee, will take subject to equities and defences of a personal character, whether it knew of the facts giving rise to such equities or defences or not.[86] The terms of the letter of credit may, however, by contract grant assignees a higher level of protection.[87]

Turning now to recent developments concerning forgery and nullity in the law of letters of credit, it is useful to start by rehearsing the relevant obligations of the banks in letter of credit operations, starting with the UCP 500 Rules because of their very frequent application to letter of credit contracts, pursuant to the intention of the contracting parties. Even as they apply, however, to a letter of credit contract governed by English law, the latter will fill in any gaps in the coverage of UCP 500. To the extent that the rules in UCP 500 are ambiguous then, subject to the following point, English law may be invoked to clarify the ambiguity. Suppose, however,

[81] *Arab Bank Ltd v Ross* [1952] 2 QB 216, 227 (Denning LJ).
[82] *Safa Ltd v Banque du Caire* [2000] 2 Lloyd's Rep 600 (CA).
[83] Bills of Exchange Act 1882, s 24.
[84] Bills of Exchange Act 1882, ss 29, 38.
[85] Bills of Exchange Act 1882, s 29(2).
[86] *Banco Santander SA v Banque Paribas* [2000] 1 All ER (Comm) 776 (CA). See also *Solo Industries UK Ltd v Canara Bank* [2001] 2 Lloyd's Rep 578, 587 (CA).
[87] *Standard Bank London Ltd v Canara Bank* (Moore-Bick J 22 May 2002); *Banco Santander SA v Bayfern Ltd* [1999] 2 All ER (Comm) 18 (Langley J).

that the International Chamber of Commerce[88] issues an opinion on the meaning of a particular provision. Doubts may be expressed as to the admissibility of this opinion in judicial proceedings[89] but probably the better view is that parties incorporating the Rules do so with any accompanying authoritative opinions issuing from the ICC. However that may be, English procedural rules, and in particular CPR (Civil Procedure Rules) Part 24,[90] will apply in respect of letter of credit proceedings in this country. Furthermore, to the extent that English law imposes mandatory rules on the letter of credit contract, these would apply even if they overrode express provisions of UCP 500. Rules relating to fraud, even though they are not legislative in character, would appear to have the quality of mandatory rules and so, even if UCP 500 did not silently defer to the fraud rules of the applicable law, should be applied in English courts on that account.

Under UCP 500, the duty owed to the beneficiary of a letter of credit by the issuing bank is a duty to make payment[91] against 'stipulated document(s)'.[92] This of course prompts the question what exactly is a stipulated document, which should be answered with due account taken of the absence in UCP 500 of any express duty on the issuing bank to the beneficiary to pay against documents that only apparently conform. Article 13 of UCP 500 also imposes a duty on the issuing bank—and on other banks taking up documents—to 'examine all documents stipulated in the Credit with reasonable care, to ascertain whether or not they appear, on their face to be in compliance with the terms and conditions of the Credit'. This is not the same as a duty to take reasonable care in determining whether the documents tendered are the stipulated documents. A bank that fails to comply with its duty of examination may not expect to be reimbursed for making payment to the beneficiary. A bank that does pay against apparently conforming documents, even if they are not the stipulated documents, is consequently entitled to reimbursement. This accords with ordinary agency principles and indeed is expressly called for in UCP 500 on the part of an issuing bank towards a nominated bank that has made payment to the beneficiary.[93] The bank is entitled to be reimbursed by its customer even if the documents it takes up are forgeries[94] and it may not be prevented from paying against documents merely because the applicant alleges that fraud has

[88] The body that promulgates and publishes UCP 500.
[89] Doubt was expressed in *Banco Santander SA v Banque Paribas* [2000] 1 All ER (Comm) 776 (CA) about the admissibility of an ICC Banking Commission Statement on the Future of UCP 500 revision.
[90] This deals with summary judgment.
[91] I shall in this paper assume that the bank undertakes to pay cash against documents, though there are various ways in which payment may be effectuated, unless otherwise indicated.
[92] Arts 2 and 9.
[93] Art 14.
[94] *Gian Singh & Co Ltd v Banque de l'Indochine* [1974] 1 WLR 1234, 1238 (Lord Diplock).

been committed.[95] Given the essentially clerical character of the bank's performance of its duty of examination, it is questionable whether there is much of a distinction between a strict duty and a duty of reasonable care to examine the documents for apparent compliance. It is hard to see how a bank can have taken care in making payment against documents that are not in apparent compliance with the credit. Whether there would be a difference if the duty related to the documents being the stipulated documents depends upon the meaning of 'stipulated document(s)', on which so far no light has been shed.

Before dealing further with this question, account should be taken of Lord Diplock's speech in *United City Merchants (Investments) Ltd v Royal Bank of Canada*,[96] where his lordship observed that it was 'trite law' that if, 'on their face, the documents presented to the confirming bank by the seller conform with the requirements of the credit as notified to him by the confirming bank', the bank would be bound to pay the seller. Since this principle is not stated in UCP 500, it calls for further elaboration, which to a very limited degree it receives in Lord Diplock's later assertion that it would be 'strange' if the duty to the seller were not to pay against 'the presentation of apparently conforming documents' when this was the duty owed by confirming bank to issuing bank and by issuing bank to applicant.[97] With respect, it would not be strange at all if the beneficiary's right to be paid were not the same as the bank's right to be reimbursed for having made payment. A letter of credit contract is not a contract of indemnity. Yet, even where the beneficiary is guilty of fraud and thus has no right to be paid at all, the bank making payment and taking reasonable care without apprehending the fraud has a right to be reimbursed. This objection apart, the failure of UCP 500 to make mention of Lord Diplock's trite legal proposition would not prevent it from being a perfectly acceptable principle of English law to be called into play to the extent that the letter of credit contract with the beneficiary is not inconsistent with it, which in effect means to the extent that it is not inconsistent with UCP 500. Nevertheless, the application of that principle to a letter of credit contract incorporating UCP 500 must not contradict the provisions of UCP 500. So it comes down to whether the UCP 500 duty to pay against the 'stipulated document(s)' is inconsistent with a duty to pay against documents that on their face are conforming documents.

[95] *Bolivinter Oil SA v Chase Manhattan Bank* [1983] 1 Lloyd's Rep 251; *Turkiye Bankasi AS v Bank of China* [1996] 2 Lloyd's Rep 611; *Montrod Ltd v Grundkotter Fleschvertriebs-GmbH* [2001] All ER (Comm) 368 (HH Judge Jack).

[96] [1983] 1 AC 168.

[97] In *Montrod Ltd v Grundkotter Fleschvertriebs-GmbH* [2001] All ER (Comm) 368, however, HH Judge Jack notes the difference in language in UCP 500 between a bank's duty to examine documents to see if on their face they appear to be compliant and its duty to the beneficiary to pay against the stipulated documents but is nevertheless persuaded that the bank should not be allowed to go behind the appearance of the documents to justify a refusal to pay.

Take the following two cases: first, the bill of lading that has been falsely dated (a forgery); and secondly, the bill of lading that represents a phantom (ie, a non-existent) shipment (a nullity). Suppose that the letter of credit calls for a bill of lading representing the shipment in March of 5,000 tonnes of barley from a South Australian port. In both cases, the presence of fraud on the part of the beneficiary will provide the bank with a defence to payment, but let us suppose that the beneficiary is an intermediate trader in string who has not been fraudulent at all. Whether the bill of lading is falsely dated on 31 July when the goods were in fact shipped on 1 August, or concerns a phantom shipment, it will appear on its face to comply with a letter of credit so that, further to Lord Diplocks's 'trite' proposition, payment has to be made under the letter of credit contract. A falsely dated bill of lading representing an actual cargo and signed by the ship's master or agent may indeed, it is submitted, be a stipulated document, since the bill of lading bears the date of a July shipment as required by the bank. This is further consistent with the decision of the House of Lords in *United City Merchants* that a defence to payment should not exist merely because the documents tendered contain a 'material statement that is inaccurate'.

The bill of lading concerning the phantom shipment, however, is a different matter. It may be on a carrier's or charterer's standard form, and it may carry the genuine signature of a dishonest ship's master or the forged signature of an honest master,[98] yet it cannot, as we have seen, be said to be a bill of lading at all and so is a nullity. It cannot be a stipulated document as required by UCP 500. As between a buyer and seller on CIF or similar terms, the duty of the seller in respect of misstatements in the documents tendered has already been noted. In the CIF case of *Hindley & Co Ltd v East Indian Produce Co Ltd*,[99] the bill of lading represented goods that had not been shipped at all, Kerr J stating the case to be 'of an *a fortiori* nature' in comparison with other cases of documentary misstatements. The bill, he said, was not a document of title at all and, drawing upon other authority, was not a 'valid and effective document'.[100] In other words, the bill of lading was a nullity in all but name. Although the law of CIF sales and of letter of credit contracts need not be in exact alignment, they ought not to differ in their treatment of nullities. In consequence, there is all the more reason not to read down the bank's duty to pay against stipulated documents so as to compel it to pay against documents that are apparently conforming when these documents are nullities.

In *United City Merchants*, the defendant confirming bank was unable to resist payment where the fraud in question was not the fraud of the beneficiary

[98] In the latter case only would it be a forgery.

[99] [1973] 2 Lloyd's Rep 515.

[100] *Ibid*, 519. See also *Rafsanjan Pistachio Producers Cooperative v Bank Leumi plc* [1992] 1 Lloyd's Rep 513, 540 (invoice recording a non-existent transaction a 'complete concoction').

but instead the fraud of a loading broker independent, and not an agent, of the beneficiary. The same distancing of the beneficiary from the source of the fraud would also arise in the common case of the beneficiary who is an intermediate string trader. Now, it is precisely because third party fraud was thus held to be no defence to payment under a letter of credit that forgery and nullity, hitherto in the shadow of fraud, came into their own as possible payment exceptions on their own account. The bank in *United City Merchants* therefore sought to establish a defence beyond fraud. Lord Diplock, while rejecting any defence to payment based upon material misstatements in the documents (which in at least some cases would amount to forgery), left open the question of the beneficiary's rights against a conforming bank where, unknown to that beneficiary, the document was a nullity because it had been forged by a third party. The issue did not arise in the present case because the bill of lading, though falsely dated, 'was far from being a nullity' and was 'a valid and transferable receipt for the goods'. In that the date of shipment was falsified, the bill of lading was a forgery, so that the rejection by the House of Lords of a payment exception in *United City Merchants* is a definitive rejection of a specific forgery exception.[101] In *Solo Industries Ltd v Canara Bank*,[102] however, Mance LJ comes close to going beyond Lord Diplock's position by recognising that forgery is an exception to the bank's duty to pay under a letter of credit contract. The example he cites, nevertheless, a document forged by the beneficiary, makes it equally an example of fraud on the part of the beneficiary.

The issue of nullity, left open by Lord Diplock, was taken up in *Montrod Ltd v Grundkotter Fleischvertriebs-GmbH*.[103] In that case, HH Judge Jack at first instance uncompromisingly rejected a separate nullity exception to the bank's duty to make payment under a letter of credit contract. He considered that such an exception ran counter to the 'fundamental principle that banks consider the documents alone and should not take account of other matters, in particular disputes between applicant and beneficiary'. With respect, given the potential scope of the fraud exception—which certainly does compromise that fundamental principle—and the suitably restricted definition of a documentary nullity, this looks rather like swallowing a camel and straining at a gnat. Furthermore, to refuse to pay against a documentary nullity is not to become embroiled in the underlying transaction. Again, the bank is asserting a position that is no different from that of the acceptor of a bill of exchange who denies liability on the ground that a signature has been forged. The document in *Montrod*, anyway, was far from being a nullity in the sense of a bill of lading representing a

[101] Cf Denning MR and Browne LJ in *Edward Owen Engineering Ltd v Barclays Bank International Ltd* [1978] QB 159, 169, 172 (bank entitled not to pay when it knows the documents are forged).
[102] [2001] All ER (D) 34.
[103] [2001] All ER (Comm) 368 (HH Judge Jack), reversed in part [2001] EWCA Civ 1954; Hooley [2002] CLJ 279.

phantom shipment. It was an inspection certificate signed by the seller of goods, when the seller lacked the authority to sign,[104] purportedly on behalf of a party (Montrod) that stood to reimburse the issuing bank in the event of that bank making payment under the letter of credit. The seller had acted, however, in good faith, having been persuaded by the buyer that it was in order for it to sign on behalf of Montrod, which had authorised the signature.[105] Consequently, the fraud exception did not apply. Furthermore, it had not forged Montrod's signature. On these particular facts, it was entirely understandable that the trial judge should not have wanted to apply any supposed extension to the fraud principle.

In the Court of Appeal in *Montrod*, the sole question was whether there should be a defence to payment in the case of nullity of documents. The court, being of the view that any nullity exception was an extension of the fraud exception, rather than something that was different in kind, concluded that no such exception should be recognised. In the judgment of Potter LJ, a general nullity exception is not 'susceptible of precision, involves making undesirable inroads into the principles of autonomy and negotiability universally recognised in relation to letter of credit transactions'.[106] The court's handling of the matter was not assisted by the way that the defence was presented: it was put in terms of the beneficiary's entitlement to payment where it was apprised of the nullity prior to payment actually being made. If there is to be a nullity exception, however, it should stand independent of the state of mind of the beneficiary, whether that state of mind is assessed prior to presentation of the documents or prior to payment by the bank or at any other time. Nullity and fraud are different, clear analysis of the issues not being assisted by invoking the pathetic fallacy of 'fraudulent documents'[107] which confuses the difference between fraud and nullity. The court, moreover, treated it as self-evident that the bank's duty to pay the beneficiary was a duty to pay against documents that on their face conformed to the letter of credit.

[104] Montrod never had any intention of inspecting the goods. The inspection clause was a 'locking in' clause, designed to allow Montrod to hold up payment until it had been put in funds to reimburse the issuer.
[105] The seller reasonably attached little significance to the inspection certificate. This was its first involvement with letters of credit. In addition, it had not been informed of the significance of the 'locking in' clause and there was to be another inspection when the meat arrived in Moscow.
[106] [2001] EWCA Civ 1954 at [58].
[107] An expression used in R Jack, A Malek and D Quest, *Documentary Credits* 3rd edn (London, Butterworths, 2001), para 9.23 ('a document may be fraudulent without being forged'). See also Potter LJ in *Montrod Ltd v Grundkotter Fleschvertriebs-GmbH* [2001] EWCA Civ 1954 at [59] ('a document … fraudulent in itself'); Browne LJ in *Edward Owen Engineering Ltd v Barclays Bank International Ltd* [1978] QB 159, ('the bank knows … the documents are … fraudulent'); Stephenson LJ and Griffiths LJ in *United City Merchants (Investments) Ltd v Royal Bank of Canada* [1981] 3 All ER 142, 164, 175 ('a document fraudulent as this one was'; 'the documents presented are fraudulently false'). This regrettable usage is also to be found in UCC Art 5–109(a) ('document … materially fraudulent').

Although a separate nullity exception under English law was rejected in *Montrod*, the starting point in any case must be the terms of the letter of credit contract itself. Just as the rights of an assignee can be amplified by the terms of a letter of credit contract, so the rights of a beneficiary may be amplified or diminished by the terms of that contract. The point was made above that the bank's duty to pay was against 'stipulated document(s)' and that UCP 500 differentiated the language of this duty from the language of the duty of other banks to indemnify the paying bank. If that language is apt to give the bank a defence against payment in the case of documentary nullities, as was argued above, then it must override any residual principle of English law if UCP 500 has been incorporated in the letter of credit contract.

Having so firmly rejected a nullity exception, pointing to the dangers it posed for the international banking system, the Court of Appeal in *Montrod* then went on to declare some sympathy for an altogether different defence to the bank's duty to pay based upon a Singaporean decision[108] to the effect that recklessness and haste on the part of a beneficiary, falling short of fraud but assisting in the perpetration of fraud, should prevent the beneficiary from making a claim under the letter of credit. This decision appears to form part of an emergent exception in that jurisdiction to the bank's duty to pay based upon unconscionable conduct on the part of the beneficiary.[109] If a nullity exception eludes definition, one wonders what to make of a very vague exception of this sort and why the court should have given such a hostage to fortune. The mischief potential of this exception to payment under a documentary credit is very considerable. This exception is likely to result in more cases where payment need not be made than any nullity exception.

There is a final and difficult point emerging from the decision of the Court of Appeal that goes to the scope of the fraud exception. It concerns an untruthful document where the beneficiary is not responsible for the untruth but is aware of it at the time of presentation.[110] The implication is that this knowledge disqualifies the beneficiary from obtaining payment. If, however, the beneficiary is not responsible for the untruth, and if there is to be no nullity extension to the fraud exception, it is difficult to see what the

[108] *Lambias (Importers and Exporters) Co PTE Ltd v Hong Kong & Shanghai Banking Corpn* [1993] 2 SLR 751.

[109] See R Jack, A Malek and D Quest, *Documentary Credits* 3rd edn (London, Butterworths, 2001), para 9.25 and decisions therein cited.

[110] Citing *Group Josie Re v Walbrook Insurance Co Ltd* [1996] 1 WLR 1152, 1161: '[I]t is nothing to the point that at the time of trial the beneficiary knows, and the bank knows, that the documents presented under the letter of credit were not truthful in a material respect. It is the time of presentation that is crucial.' See also Ackner LJ in *United City Merchants (Investments) Ltd v Royal Bank of Canada* [1981] 3 All ER 142, 170: 'If the signature on the bill of lading had been forged, a fact of which the sellers were ex hypothesi ignorant, but of which the bank were aware when the document was presented, I can see no valid basis on which the bank would be entitled to take up the drafts and debit their customer'. This type of forgery, nevertheless, would also amount to a nullity.

beneficiary's knowledge of the untruth has to do with the matter. If indeed, as the court says, it is the bank's duty to pay against apparently conforming documents, and if the beneficiary has not been guilty of fraud, then why should the beneficiary's knowledge matter, whether or not the beneficiary keeps quiet about the documentary misstatement? Test it in this way. Suppose that the beneficiary is a party in the middle of a sales string and accepts the documents from a prior seller before discovering the misstatement. The beneficiary then, in the spirit of utmost candour, informs the confirming bank of the untruth but reminds the bank at the same time that it remains bound to pay. What possible defence does the bank have to payment, unless an unconscionability exception of some sort is adopted? Suppose now that the beneficiary remains silent. What difference should that make? The letter of credit contract has long since been concluded and any implied misrepresentation that might arise from the silent presentation of the documents certainly does not induce the letter of credit contract or affect payment. Besides, even if there were such a misrepresentation, in what sense could it be material, given that the bank was bound to pay anyway? And why should it matter whether the beneficiary's knowledge arises before presentation or before payment? To assist clear reasoning in the matter, it is necessary to avoid the imprecise language of 'fraudulent documents'. If the candid beneficiary is to be disqualified in some cases at least, it must be because the beneficiary is tendering a documentary nullity.

The arguments advanced above would, if adopted, in one respect diminish the gap between the treatment of documents tendered under a CIF contract and under a letter of credit contract and in another respect widen it. They would diminish the gap to the extent that, just as a CIF seller may reject documentary nullities, so should a bank be entitled to reject similar documents under a letter of credit contract. The gap would not be eliminated because forged documents, if not nullities, would in the absence of fraud be stipulated documents for the purpose of payment under the letter of credit, though they would not constitute good tender under a CIF contract. The above arguments would widen the gap to the extent that Lord Diplock's dictum, that a CIF seller is bound to pay against documents that on their face are conforming, is regarded as accurately stating the present law, which it does not.

IV. CONCLUSION

The argument advanced in this chapter is that the documentary requirements that have to be satisfied under a letter of credit contract are not the same requirements as those resting on a seller under a CIF contract. The CIF seller must tender documents that do not *in fact* contain misstatements. Under a letter of credit contract, the orthodox view is that a bank is bound to pay against *apparently* conforming documents, though a closer reading of UCP 500 with its reference to 'stipulated document(s)' would narrow the gap between CIF and letter of credit contracts. A fraud exception is not needed to qualify the CIF seller's duty; it has however been recognised as qualifying the bank's payment duty under a letter of credit contract. It has also been argued in this paper that, notwithstanding contrary authority, a separate exception to the bank's duty to pay should be recognised in the case of documentary nullities, a category that is far from comprising all forgeries and all documents containing false statements.

That the documentary requirements of the two types of contract should be different—notwithstanding Lord Diplock's attempt to restate them in identical terms—is no cause for reproach. They are very different contracts. The performance of CIF contracts is by no means the same thing as the performance of financial obligations resting against the backdrop of a banking system and involving instruments that to a marked extent are approximated to negotiable instruments. The autonomy principle is needed to facilitate the flow of payment in just the same way as a bill of exchange derives its character from being an unconditional order in writing divorced from the underlying transaction.[111] There is to a significantly lesser degree acceptance of autonomy within the CIF contract in terms of the separation of its physical and documentary aspects. The *Panchaud Frères* principle, for example, is flatly inconsistent with autonomy in this sense.

Nevertheless, whereas types of contract considered individually may be subjected to their own internal logic, it may be seen as awkward that contracts that inhabit the same network of contracts should jar when placed side by side. Although cash against documents may be the payment system of choice in the case of dry cargoes, letters of credit are common forms of payment in contracts for the sale of oil. If a CIF seller's right of payment is not the same under the contract of sale as it is under the letter of credit contract, then a way has to be found to reconcile the inconsistency. The relationship of inconsistent FOB and letter of credit contracts was considered at length by Robert Goff J in *Ficom SA v Sociedad Cadex Lda*,[112] where three possibilities were recognised. First, where buyer and seller subsequently agree on the terms of the letter of credit contract, this may amount to a variation of the sale contract. Secondly, if the contract of sale does not define the terms of the letter of credit, the parties' subsequent agreement on the terms of the letter of credit contract may also supplement the terms of the sale contract. Thirdly, where the letter of credit contract departs from the terms of the sale contract, there may be conduct on the part of the seller that amounts to a forbearance 'giving rise at most to an equitable estoppel, and may be capable of withdrawal … upon … reasonable notice'.

[111] Bills of Exchange Act 1882, s 3.
[112] [1980] 2 Lloyd's Rep 118, 131 (Robert Goff J).

The analysis in *Ficom* may suit cases where the contract stipulates one type of letter of credit but the applicant provides another, or the documents listed in the sale of goods contract are not the same as those listed in the letter of credit contract. It does not, however, fit the case of different standards of documentary compliance as between CIF and letter of credit contracts. The opening of a letter of credit amounts only to conditional payment.[113] If payment is not in fact made, then the seller's right to be paid according to the terms of the sale contract, previously held in suspense, is revived. If payment is made against apparently conforming documents, with the result that the buyer has to reimburse the issuing bank, but the buyer then goes on to reject the goods for late shipment, the seller may not retain the moneys. Although the seller will have been paid by the bank and not the buyer, the seller will come under an implied contractual duty to the buyer to account for the money received.[114] The letter of credit contract gives the seller beneficiary a right to be paid and not a right as against the buyer to retain the moneys when this is not in accordance with its rights under the sale contract. Although the CIF buyer's and the bank's duties of payment are different, there is in the end no inconsistency between the two contracts.

[113] *WJ Alan & Co Ltd v El Nasr Export and Import Co* [1972] 2 QB 189.
[114] *Cargill International SA v Bangladesh Sugar & Food Industries Corpn* [1996] 4 All ER 563, affd [1998] 1 WLR 461 (performance bond).

Commentary on 'Documents and Contractual Congruence in International Trade'

WILLIAM BLAIR, QC

IN HIS SUPERBLY lucid chapter, Professor Michael Bridge discusses several issues that arise out of the use of documents in international trade transactions. He considers first the position between buyer and seller under a CIF contract, that is, a contract for sale the price of which includes the cost of shipment and insurance. It is performed by the seller tendering shipping documents to the buyer that conform to the contractual requirements. Professor Bridge demonstrates that, notwithstanding certain dicta in *Gill & Duffus SA v Berger & Co Inc*[1] that the obligation is to tender documents which on their face conform to those called for by the contract, the correct view is that if the documents are not genuine, the buyer has the right to reject them even if the forgery is not apparent. The second situation he considers, and the one on which I would like to comment, concerns the financing of such a transaction by a documentary credit. In this case, the insertion of one or more banks into the process introduces further contractual relationships.

When a sales transaction is financed by a documentary credit, the buyer (applicant) requests a bank in his own country to issue the credit in favour of the seller (the beneficiary) in his country. The buyer specifies the documents against which the bank should pay. The rule is that documents tendered must strictly comply with the requirements of the credit. The buyer's bank in turn usually instructs a bank in the seller's country to advise the seller of the credit. That bank may act as negotiating bank, and sometimes adds its own confirmation to the credit. Upon shipment, the seller presents the documents to the advising bank, which makes payment if the documents are in order. It is reimbursed by the issuing bank, which in turn is

[1] [1984] AC 382.

reimbursed by the buyer. When the documents are presented to the bank by the seller the examination is, as Professor Bridge points out, largely clerical. When, as frequently occurs, there is some discrepancy between the documents as tendered and as required by the terms of the credit, the bank concerned may ask the buyer if it wishes to waive the discrepancies. The whole process has to be carried out within a short timescale. Article 13(b) of the Uniform Customs and Practice for Documentary Credits (UCP 500) provides for a maximum of seven banking days following the receipt of the documents. Within that time, the bank has to examine the documents and decide whether to take them up or to refuse them.

As regards the nature of the bank's duty, Professor Bridge points out that a number of provisions in the UCP are relevant. The bank undertakes to pay provided that the stipulated documents are presented (Articles 2 and 9). The letter of credit contract is 'separate' from the underlying contract, the bank in no way being 'concerned with or bound' by such contract (Article 3). This is a statement of the autonomy principle, which is also recognised at common law. The parties to credit operations deal with documents and not with the goods or services to which the documents relate (Article 4). A key provision is Article 13a, which stipulates that 'banks must examine all documents stipulated in the credit with reasonable care, to ascertain whether or not they appear, on their face, to be in compliance with the terms and conditions of the credit'. If the bank satisfies that obligation, it is entitled to reimbursement even if the documents subsequently turn out to be forged.[2]

However the question of forged documents and fraud generally raise some tricky questions, particularly where there are a number of banks in the chain. Whilst the UCP itself does not deal with the issue, under English law (as in many other jurisdictions) the principle is that, even if the documents are on their face compliant, the bank must not pay in cases of 'established fraud'. If in the rare case in which fraud is in this sense established (as opposed to being alleged, which is a much more common but very different thing) the bank nevertheless pays, it loses its right to reimbursement. But the precise scope of the 'fraud exception' has been the subject of argument.

The leading case is *United City Merchants (Investments) Ltd v Royal Bank of Canada*.[3] In that case, the date of shipment was falsely stated in the bill of lading as being the last day for shipment provided for by the credit, whereas the goods had in fact been shipped on the following day. The false date had been inserted by a third party, namely the loading brokers. The sellers (beneficiaries) were unaware of the true position when they presented the documents. It was held that the issuing bank was obliged to pay. Lord Diplock (who gave the leading judgment) defined the fraud

[2] *Gian Singh & Co Ltd v Banque de l'Indochine* [1974] 1 WLR 1234.
[3] [1983] 1 AC 168.

exception as applying to fraud on the part of the beneficiary. He did however leave open the question of the rights of an innocent seller/beneficiary against the bank when a document presented by him is a nullity because unknown to him it has been forged by some third party. That issue did not arise for consideration in *United City Merchants*, because the view was expressed (correctly) that the mere fact that a bill of lading bears the wrong date does not make it a nullity.

In *Montrod Limited v Grundkötter Fleischvertriebs GmbH*,[4] the nullity issue was directly addressed, though the facts were different from those postulated by Lord Diplock, because the documents alleged to be nullities were created by the beneficiary, not a dishonest third party. In that case, the terms of the credit called for inspection certificates. Those tendered were signed by the beneficiary in the applicant's name. The trial judge concluded that this was done in the honest but mistaken belief that it was authorised by the applicant, so that there was no question of fraud on the part of the beneficiary.[5] The certificates were in apparent compliance with the requirements of the credit, and the bank paid. It was held to be entitled to be reimbursed by the applicant. An argument that, independent of the fraud exception, a bank is not obliged to pay against a document which though apparently complying, is in fact a nullity, was rejected both at first instance,[6] and by the Court of Appeal. The latter stated that:

> The combination of the autonomy principle and the rule that the banks concerned deal in documents and not in goods (articles 3 and 4), together with the issuing bank's undertaking of payment if the stipulated documents presented conform with the terms of the credit (see article 9), plainly entitled [the] beneficiary to obtain, and obliged [the] issuing bank to make, payment against the documents presented provided that they complied 'on their face' with the requirements of the credit: see articles 13(a), 14(a), 14(b) and 14(c). (Paragraph 37)

Any 'nullity exception' additional to the fraud exception was rejected on policy grounds, the Court saying that:

> The argument for *Montrod* that, where fraud on the part of the beneficiary cannot be established, there should nonetheless be room for a nullity exception in the case of a document which is worthless in the sense that it is not genuine and has no commercial value, whether as a security for the goods or otherwise, involves an undoubted extension of the fraud exception as hitherto propounded in the English authorities. If the basis of a fraud exception is that

[4] [2002] 1 WLR 1975.
[5] An earlier application by the beneficiary for an injunction based on the fraud exception had been dismissed by David Steel J.
[6] 28 November 2000, Judge Raymond Jack QC (now Jack J), who is co-author of the leading textbook, Jack, Malek & Quest, *Documentary Credits*, 3rd edn, (London, Butterworths, 2001).

> the court will only intervene in breach of the autonomy principle for the purpose of preventing or discouraging the perpetration of fraud on the part of the beneficiary or other presenting party, it is a clear extension to hold that presentation of a document which is itself a nullity for reasons which are *not* known to the beneficiary or issuing bank at the time of presentation, are nonetheless to be similarly treated. (Paragraph 43)

On 30 May 2002, the Appeal Committee of the House of Lords dismissed a petition by Montrod (the beneficiary) for leave to appeal. Thus the position for the time being must be taken as settled.

Standing back and looking at the issue from a policy perspective, there is no debate as to the strict limitations that must be placed on the circumstances in which a bank is entitled (or obliged) to refuse payment when allegations of fraud, documentary nullity and the like are raised. The buyer should not be permitted to hold up payment because of complaints as to performance of the sale contract. The position is best demonstrated by reference to performance guarantees, which are analogous instruments, but usually of much longer duration. The applicant may have sound grounds for believing that complaints as to performance giving rise to a call on the guarantee are in fact unjustified. (Contrast the case of a documentary credit, in which payment will probably be made before the goods arrive at their destination, and before the buyer becomes aware of the state of the goods). Nevertheless, absent fraud, the bank must pay. As Professor Goode has pointed out, one of the primary functions of these instruments is to create an abstract payment obligation independent of and detached from the underlying contract.[7] As such, these instruments play an important role in international trade finance. There are reputational implications for an institution that refuses payment.

However there remains a legitimate question as to whether the exception as defined in *United City Merchants* is too narrow. Professor Bridge powerfully argues that notwithstanding this authority as confirmed by *Montrod*, a separate exception to the bank's duty to pay should be recognised in the case of documentary nullities. He emphasises that this is not a category that comprises all documents containing false statements. The paradigm is a false document issued as regards a non-existent cargo. Plainly such a document is a nullity. He argues that the bank's obligation is to pay against the stipulated, not apparently conforming, documents. The effect of UCP Article 13 is that where it has without negligence paid on documents which are on their face compliant, a bank is entitled to reimbursement, but it does not follow that the bank's obligation is to pay against apparently conforming documents. He concludes that in the exceptional case in which the documents tendered are in fact nullities—for example, the phantom cargo situation—the

[7] *Commercial Law*, 2nd edn (London, Penguin, 1995) p 987.

bank cannot be required to pay, regardless of the honesty or otherwise of the beneficiary.

The answer to the question can best be tested against the likely practical consequences, taking to begin with Professor Bridge's example of a beneficiary who honestly takes apparently conforming documents which are in fact nullities, and discovers their falsity prior to presentation to the advising bank. What is the legal position in that case? If the beneficiary goes ahead and presents the documents to the bank, it is bound to pay? In my view, the answer is that it is clearly not, but I do not think that there is any need to have regard to a 'nullity exception' to reach that result. The honesty of the beneficiary is to be judged at the time of presentation of documents.[8] A beneficiary cannot honestly present documents he knows to be forged to a bank for payment, even if he took the documents innocently. The fraud exception applies to such a case. Further, it would seem reasonable to infer that in presenting documents to the bank, the beneficiary impliedly represents that he believes them to be genuine.

Whilst that answers one point raised by Professor Bridge, there is another factual situation which is more difficult to analyse in terms of the fraud exception. This is the case where the beneficiary remains unaware of the forgery at the time of presentation, so that he is at all times honest, but the bank discovers the true position prior to payment. Is it obliged to pay if the documents are on their face compliant? This is the effect of *United City Merchants* and *Montrod*, but it would seem odd if the bank were required to pay in such circumstances. A 'nullity exception' could provide an answer in such a situation.

But is it necessary or desirable? Again, there are important practical considerations. As has been seen, the bank has a maximum of seven working days in which to take up the documents, or reject them. Under the law as it stands, and leaving aside the possibility of an injunction, the bank must pay unless within that time fraud on the beneficiary's part is established. This is a high hurdle to surmount. One could say that nullity would have to be established to a similar degree of proof, so that nothing is lost by introducing such an exception. On the other hand, nullity is a more technical concept than fraud. There is an interesting discussion in Professor Bridge's chapter in this regard, in which a distinction is drawn between forgeries and nullities, and documents which albeit containing misrepresentations fall into neither category. One can say clearly that documents issued in connection with a phantom cargo are nullities, but in other cases the line may be harder to draw. It is also worth making the point that parties may be more willing to allege nullity than fraud, and banks may find the allegation less easy to evaluate within the confines of their Article 13 duties, though doubtless the last point is debatable.

[8] *Group Josie Re v Walbrook Insurance Co Ltd* [1996] 1 WLR 1152 at 1161.

However the possibility remains of *incontrovertible* evidence of fraud emerging after an honest presentation by the beneficiary, but before payment by the bank. If the shipment never took place, there is no room for argument that the documents are other than nullities. Whether or not they appear plausible on their face, the documents are worse than waste paper: they are implements of fraud. If one assumes that the seller presented the documents not knowing or suspecting the true facts, the fraud exception as defined in *United City Merchants* does not apply.

Nonetheless, particularly bearing in mind that a decision must be taken within seven working days, common sense suggests that it is most unlikely that a bank would pay in the face of incontrovertible evidence. If it did, then sought to pass the documents on for reimbursement, payment would likely be refused by the other banks in the chain, as well as by the applicant. The most likely ground for refusal would be established fraud, the point being that in exceptional circumstances like these, it would not be sufficient for the beneficiary to assert his innocence. The evidential burden would be on him to prove it.

Whilst this may be a likely result in practice, it does not answer the point of legal principle. Suppose that the beneficiary is able to prove his innocence at time of presentation, and sues the bank for non-payment? A possible answer is a procedural one, and here another important point made by Professor Bridge comes into play. That concerns the position as between buyer and seller. If the bank refuses to pay, the seller's right to be paid according to the terms of the sale contract, previously held in suspense, is revived—but there will no such right if the goods were never shipped. On the contrary, the seller will be in breach of the contract of sale. In the hypothetical legal proceedings brought by the seller/beneficiary, the bank will join its correspondent bank to the action, and it will in turn join the applicant. The applicant/buyer will claim damages against the seller for breach of contract, so that it is hard to see the advantage to the seller in suing.

It is also interesting to consider the converse situation in which payment against apparently conforming documents *is* in fact made by the bank. If the goods were never shipped, Professor Bridge points out that the seller will come under an implied contractual duty to the buyer to account for the money received.[9] Thus it appears that where nullity arises from non-shipment, the loss is likely in the result to fall on the seller, whether he is honest or not, and without needing to introduce a nullity exception into the law.

Finally, Professor Bridge takes issue with another aspect of *Montrod*. The Court of Appeal left open the possibility of an 'unconscionability exception' to the bank's duty to pay based on a decision of the High Court of Singapore to the effect that recklessness and haste on the part of a beneficiary,

[9] See *Cargill International SA v Bangladesh Sugar & Food Industries Corpn* [1996] 4 All ER 563, affd [1998] 1 WLR 461 (a case involving a performance guarantee).

falling short of fraud but assisting in the perpetration of fraud, should prevent the beneficiary from making a claim under the letter of credit.[10] He says tellingly that 'if a nullity exception eludes definition, one wonders what to make of a very vague exception of this sort and why the court should have given such a hostage to fortune. The mischief potential of this exception to payment under a documentary credit is very considerable. This exception is likely to result in more cases where payment need not be made than any nullity exception.' An 'unconscionability exception' is also rejected by Jack,[11] in my view correctly.

[10] *Lambias (Importers and Exporters) Co PTE Ltd v Hong Kong & Shanghai Banking Corpn* [1993] 2 SLR 751.
[11] See Jack, Malek & Quest, *Documentary Credits* (*ibid*) at para 9.25.

10

The Dematerialisation of Money Market Instruments

JOANNA BENJAMIN[1]

PART I

I. OVERVIEW

THIS CHAPTER CONCERNS the impact of electronic technology on an area of commercial law and practice. It considers the dematerialisation of money market instruments ('MMIs'). A number of important papers have been written on this and related topics.[2]

MMIs are highly liquid,[3] short-term debt securities[4] such as certificates of deposit, commercial paper and certain bills of exchange. In their

[1] The author is grateful to Guy Morton of Freshfields who commented on an earlier draft of this chapter at the Commercial Law and Commercial Practice Seminar, London School of Economics, 29–30 November 2002. His comments were invaluable.

[2] These include the following: E Micheler, 'Farewell Quasi-Negotiability? Legal Title and Transfer of Shares in a Paperless World' (2002) *JBL* 358–378; Bank of England, 'The Future of Money Market Instruments' (1999) (written by Guy Morton, partner, Freshfields); JS Rogers, 'Policy Perspectives in the Revised UCC Art 8' (1996) 6 *UCLALR* 1413; JL Schroeder, 'Chix Nix Bundle-O-Stix: a Feminist Critique of the Disaggregation of Property' (1994) *Michigan LR* 92 239; CW Mooney, 'Beyond Negotiability' (1990) 12 *Cardozo LR* 305; W Blair, 'Negotiability and Estoppel' [1988] *The Company Lawyer* 8; JS Ewart, 'Negotiability and Estoppel' (1900) 16 *LQR* 135.

[3] Very broadly speaking, an asset is liquid if there is a ready market for it, so that it may be bought and sold without delay at a fair price. However, '[I]t is true that liquidity is a slippery concept…' R Lee, *What is an Exchange? The Automation, Management, and Regulation of Financial Markets*, (Oxford University Press, 1998) 292.

[4] In 2000 a debate took place in the London legal community as to whether MMIs satisfied the definition of securities for the purposes of the section 207 of the Companies Act 1989, which provides in primary legislation the statutory basis for CREST. It is understood that leading counsel expressed doubt that MMIs were securities, as payment obligations under them were not secured. Originally, the term 'securities' denoted security interests (such as mortgages and charges) supporting the payment of a debt or other obligation. However, HM Treasury expressed their view, in a letter to CRESTCo, that this original connotation has now been lost, and that '… we do not consider there is any requirement for a security to confer a proprietary interest in the fund or assets to which it relates' (letter, Dilwyn Griffiths, HM Treasury, to Iain Saville, CRESTCo, 19 July 2000).

traditional form, MMIs are legally categorised as negotiable instruments. They are widely used in the wholesale financial markets as alternatives to cash.[5] Dematerialisation is the process of replacing paper instruments with electronic records.[6]

A merger of two electronic securities settlement (ie delivery) systems operating in London, namely the Central Moneymarkets Office ('CMO') and CREST (which settles equities and other registered securities), is proposed to take place in 2003. The merger will complete the process of the dematerialisation of MMIs in London. The nature of traditional MMIs, their negotiable status and the process of dematerialisation are considered in the first part of this chapter.

Considerable work has been undertaken to assess the commercial, legal and operational impact of the proposed merger.[7] The key operational result of dematerialisation will be to render MMIs intangible and fungible.[8] The impact of this on negotiable status will be considered in Part II of this chapter.

> Negotiability confers property rights on holders which are superior to those conferred only by contract, in particular the right to take title free from any preceding equities.[9]

It will be argued that intangibles and fungibles may be subject to property rights, and that fungibles (although not intangibles) may probably be subject to possession. It will be further argued, however, that intangibles and fungibles cannot be negotiable instruments.

Part III of the chapter will go on to consider the consequences of loss of negotiable status. It will argue that the chief issue is security of transfer, and conclude that in general this will not be reduced. Through detailed industry consultation and the use of specialised working groups, the Bank of England and CRESTCo[10] have sought and received clear industry support for the initiative.[11] In relation to security and evidential issues, CREST is

[5] When Barings Bank failed in 1995, many pension funds and other custody clients became averse to the credit risk involved in leaving large sums of money on deposit with custodian banks. In response to these concerns, there were developed in London US-style 'cash sweep' arrangements, in which surplus cash balances were automatically transferred into 'cash funds'. These are collective investment schemes whose portfolios are invested in MMIs. MMIs are colloquially referred to as 'near cash'.

[6] Dematerialisation is defined in *Clearance and Settlement in the World's Securities Markets*, Group of Thirty, 1989, as 'The elimination of physical certificates or documents of title which represent ownership of securities so that securities exist only as computer records.'

[7] For relevant publications, see *inter alia* Bank of England, n 2; 'The Future of Money Market Instruments: Next Steps', Bank of England, March 2000; and 'Money Market Instruments in CREST—Consultation Response', CRESTCo, October 2002.

[8] As discussed in detail in Section V, 'fungible' means, broadly, interchangeable.

[9] CRESTCo, n 7, 5.

[10] CRESTCo is the operator of CREST.

[11] See the summary on http://www.bankofengland.co.uk/markets/money/mmfuture.htm.

supported by robust legal[12] and operational systems, so that the records of CREST are understood to be less prone to fraud, error or destruction than their paper equivalent. Thus, it is argued, dematerialisation will safeguard existing rights, and not significantly alter the position of the parties involved.[13]

However, lawyers in London and other common law jurisdictions have expressed concern about the dematerialisation of MMIs. The concern appears to derive from a general conceptual difficulty rather than any particular legal or operational problems. In an attempt to explain this, Part IV of the chapter will analyse the wider conceptual implications of dematerialisation. It will be argued that the dematerialisation of MMIs has latent semiotic, philosophical, psychoanalytic and religious associations, and further that these associations may explain part of the conceptual resistance that dematerialisation often encounters among common lawyers.[14] This wider

[12] See in particular regulation 33 of the Uncertificated Securities Regulations 2001, discussed infra in Section VI. D.

[13] Indeed, there is a statutory requirement that this should be so. S 207(4) of the Companies Act 1989 provides that the regulations permitting the dematerialisation of securities made under s 207(1) (which provides the statutory basis for CREST) '… shall be framed so as to secure that the rights and obligations in relation to securities dealt with under the new procedures correspond, so far as practicable, with those which would arise apart from any regulations under this section.'

[14] Such interdisciplinary study is useful in all branches of law. Indeed, Professor Lacey argues, in 'Philosophical Foundations of the Common Law: Social not Metaphysical', (in Horder (ed) *Oxford Essays in Jurisprudence*, Fourth Series, (Oxford University Press, 2000), that pure doctrinal legal analysis is not meaningful without social analysis.

PW Kahn argues, in *The Cultural Study of Law; Reconstructing Legal Scholarship* (University of Chicago Press, 1999) 27, that: 'We cannot grasp the law as an object of study if the conceptual tools we bring to the inquiry are nothing but the self-replication of legal practice itself.' He laments (7) '… the collapse of the distinction between the inquiring subject and legal practice …' and proposes (91) '… a new object for the discipline of law: not legal rules, but the imagination as it constructs a world of legal meaning.'

It is argued here that financial law (together with, possibly, the law of intellectual property) repays such study particularly well, for the following reason. Of course, the relationship between all branches of law and empirical fact is complex. However, '… some legal concepts apparently have more straightforward counterparts in the world of fact than do others …' Lacey, *op cit*, paraphrasing Hart. Certain branches of law (such as family law and criminal law) operate very broadly as second order analyses of factual behaviour. The narrative of each of these areas of law has a subject-matter that lies outside of itself, however complex the relationship between the narrative and the subject-matter may be. Parenting and killing are not merely legal ideas. In contrast, it is at least arguable that the whole of the subject-matter of financial law is internal to financial law. Debt, set-off, insolvency, the incorporated company—these are, arguably, all legal ideas having no non-legal counterpart. See *Meridian Global Funds Management Asia Ltd v Securities Commission* [1995] 3 All ER 918, 922 per Lord Hoffmann: 'There is in fact no such thing as the company as such, no 'ding an sich', only the applicable rules. To say that a company cannot do something means only that there is no one whose doing of that act would, under the applicable rules of attribution, count as an act of the company.' See also Blackstone, *Commentaries on the Laws of England* (facsimile of 1st edn, Chicago, University of Chicago Press, 1979), I, 464: 'It must always appear by attorney; for it cannot appear in person, being, as Sir Edward Coke says, invisible, and existing only in intendment and consideration of law.' (Of course, extensive counterarguments to this view are also available in legal scholarship.) Nevertheless, assuming the truth of Lord Hoffmann's and conceptual discussion is necessarily too brief to make any contribution to critical theory. It is hoped however that it may shed some light on the specialised area of finance which is the subject of this paper.

II. NEGOTIABLE INSTRUMENTS

A. Nature of Negotiable Instruments

MMIs in their traditional paper form have the legal status of negotiable instruments.[15] Negotiable instruments form a category of financial instrument enjoying special status under commercial law. Negotiability confers a number of possible advantages on the holder of an instrument.[16] The chief of these is security of transfer. A transfer is secure if a good faith purchaser takes the transferred asset free from adverse claims.[17] The holder in due course[18] of a negotiable instrument is not subject to adverse claims, whether from the issuer of the instrument, a predecessor in title or another third party, and can acquire good title from a thief.[19]

B. Sources of Negotiability

Not all instruments are negotiable instruments.[20] An instrument may acquire negotiable status by one of two means only. These are by statute[21] or by commercial custom.[22]

Blackstone's positions, one can argue that financial law is an exquisitely developed conceptual system that purports to refer to, and derive from, nothing beyond itself. In contrast to an autopeoitic system, it is both normatively and cognitively closed (see Black, 'Constitutionalising Self-Regulation', (1996) 59 *MLR* 24 at 44).

[15] Certificates of deposit: *Customs and Excise Comrs v Guy Butler (International) Ltd* [1977] QB 377 at 382; *Libyan Arab Foreign Bank v Bankers Trust Co* [1988] 1 Lloyd's Rep 259 at 276. The negotiable status of bills of exchange is confirmed by the Bills of Exchange Act 1882.
[16] Bank of England, n 2, Appendix II, p 12: 'Negotiable instruments have four main legal features: they can be transferred solely by physical delivery, they provide the holder with unchallengeable title, they provide certain benefits in the event of a court action, and they may include rights against two or more parties'.
[17] Adverse claims are discussed in Section VI on security of transfer.
[18] Broadly, the good faith purchaser. See the definition in s 29(1) of the Bills of Exchange Act 1882.
[19] However, these rules apply to genuine instruments. The risk of forgery remains.
[20] Famously, a bill of lading is not a negotiable instrument, although it shares certain characteristics of a negotiable instrument.
[21] Most importantly, bills of exchange, promissory notes and cheques enjoy negotiable status under the Bills of Exchange Act 1882.
[22] For example, the negotiable status of secured bearer bonds on the basis of commercial custom was recognised in *Webb v Herne Bay* (1870) LR 5 QB 642.

III. DEMATERIALISATION

The purpose of dematerialisation is to facilitate electronic settlement.

A. Electronic Settlement

Settlement means delivery. When securities are agreed to be issued to their first holders, or to be transferred from person to person (whether pursuant to a sale, as collateral for a loan or otherwise), it is necessary to arrange for the delivery of title to them. Delivery is often made against payment.

MMIs in their traditional form are bearer instruments. This means that title to them is determined by possession, and transferred by delivery (ie transfer of possession).[23] Possession means, very broadly, physical control.[24] The simplest traditional method of MMI settlement is the physical delivery from person to person of the paper instruments which constitute the MMIs, with any associated payment being made by cheque.

However, such paper-based settlement is slow, expensive and insecure. For this reason, the securities markets have, from the last quarter of the 20th century, been making the transition from paper-based to electronic settlement. Electronic settlement involves the use of settlement systems. These are systems comprising hardware and software with associated procedures and legal provisions that are designed to facilitate the electronic settlement of securities.[25] With electronic settlement, title to securities is conveniently transferred by debiting and crediting electronic accounts maintained by the settlement system in the names of the transferor and the transferee. Such transfers between accounts are known as 'book entry transfers'. An added advantage is that any associated payment can be synchronised with the transfer of title, by the simultaneous crediting and debiting of cash accounts maintained in the parties' names.[26] (The parties for whom such accounts are maintained will be referred to as the members of the settlement system.)

There are two broad models of electronic settlement, namely immobilisation and dematerialisation. With immobilisation, securities in

[23] See the Bills of Exchange Act 1882 s 2. In some cases, endorsement (ie signing on the back) and delivery are required.
[24] Possession involves (as a question of fact) physical control (actual, or symbolic, or through an agent), together with an intention to possess. See Pollock *An Essay on Possession in the Common Law*, (Clarendon Press, Oxford 1888).
[25] Settlement systems are usually operated by a financial institution or special purpose vehicle (ie, legal entity established for the purpose of operating the settlement system), and rely on a network of contracts between the operator, members, issuers and others. Many settlement systems have legislative backing.
[26] Such synchronisation is known as delivery versus payment or DVP. See the discussion of different models of DVP in *Delivery v Payment in Securities Settlement Systems*, Bank for International Settlements, 1992.

their traditional paper form are placed[27] with a bank or other financial institution acting as a depositary. The depositary holds the securities for members[28] of an associated settlement system. Title to the securities is transferred electronically across the accounts maintained by the settlement system for members. Thus, paper securities exist, but do not move from the hands of the depositary.

With dematerialisation, no paper securities exist. The securities are issued 'straight onto the screen'. Under the terms of issue of the securities,[29] credit entries in the accounts maintained by the settlement system provide the primary evidence of their existence, and determine title to them.

B. Immobilisation in the Central Money Markets Office

The CMO was established by the Bank of England in 1990 to permit the electronic settlement of MMIs. The Bank of England was the original operator of the CMO. In 1998, the Bank of England's *Securities Settlement Priorities Review* recommended that UK settlement be consolidated. In response to this, CRESTCo, the operator of CREST, took over the operation of the CMO in 1999 as an interim step towards the merger of the two systems.[30]

Originally all MMIs settled through the CMO were immobilised, and many remain so at the time of writing. Paper instruments are held by the Bank of England as depositary. It is provided that the Bank of England holds these instruments as bailee[31] for the members to whose accounts they are credited.[32] For these instruments, the effect of a book entry transfer within the CMO is an attornment[33] by the Bank of England to the transferee.[34]

As discussed in Section IIICa bearer securities are, by a legal fiction, treated as being constituted by the paper instruments that are issued in respect of them. On this basis, MMIs immobilised in the CMO are tangible assets.

[27] This means that bearer securities are held in the physical possession of the depositary. Registered securities (ie, securities title to which is determined by a register maintained by or on behalf of the issuer) are registered in the name of the depositary or its nominee.
[28] In some cases (such as Euroclear, a major international settlement system operating in Brussels), the depositary holds the assets for the operator of the settlement system, which in turn holds its interest for members.
[29] And in accordance with the terms of legislation permitting dematerialisation. In the UK, see The Uncertificated Securities Regulations 2001 SI 2001/3755.
[30] At the same time CRESTCo took over the operation of the Central Gilts Office (CGO). The CGO merged with CREST in 2000, as discussed in Section IIID.
[31] Bailment is a common law relationship of safekeeping based on the transfer of possession. See *Coggs v Bernard* (1703) 1 Comm 133.
[32] CMO Reference Manual s H.1.2, s K.2.5. See also form of Deed Pool as set out in section H (entered into by the Bank of England on 20 September 1999).
[33] Attornment is an undertaking whereby a bailee acknowledges that it holds property for a person other than the original bailor.
[34] See form of Deed Poll (Clause 2.1) and CMO Reference Manual s I.5.3. and s R.3.11.

In commercial practice, assets may be classed as fungible or non-fungible. A fungible asset is one in respect of which the customary redelivery[35] obligation is equivalent and not in specie.[36] The meaning of this is best illustrated by example. Cash is fungible[37] because, when A lends £5 to B, A does not expect the identical five pound note to be returned. She will treat the loan as discharged by the delivery of five one pound coins. However, if A lends B an antique coin, she expects the identical coin to be returned, and therefore the antique coin is non-fungible. MMIs immobilised in the CMO are non-fungible, in the following sense. When instruments are lodged at the CMO, each is mechanically stamped with a distinctive number. The MMIs of each member are held separately from the MMIs of each other member, and the distinctive number of each transferred instrument is quoted in every transfer.[38]

C. Dematerialisation in the Central Money Markets Office

Since 1994 it has been possible for sterling certificates of deposit ('CDs') to be issued into the CMO in dematerialised form.[39] In respect of dematerialised CDs, provisions[40] have been drafted with a view to approximating contractually the incidents of negotiability. These protect members to whose accounts dematerialised CDs are credited from the adverse claims of issuers[41] or other members.[42] However, the contractual approximation of negotiability is necessarily incomplete. This is because it is not possible contractually to bind a third party.[43] For this reason, the following problem arises. The MMIs held in the CMO are not always beneficially owned by members.

[35] Of course, since an equivalent asset and not the original asset in specie is received by the counterparty at the end of the transaction, arguably the obligation relates not to *re*delivery but to delivery.

[36] 'In short, the obligation is to deliver or transfer not an identified asset but anything which corresponds to the contract description.' Goode, 'Are Intangible Assets Fungible?', in Birks and Pretto (eds), *Themes in Comparative Law in Honour of Bernard Rudden*, (Oxford University Press, 2002), 101.

[37] However, see Goode, n 36, 101.

[38] MMIs in the CMO are non-fungible because (firstly) each instrument is not identical with each other and (secondly) the CMO system does not treat them as interchangeable. Of course, these are two separate points. As Goode points out (n 36), fungibility is a function of whether units are treated as interchangeable. Of course, certain types of asset (such as bank notes) may be non-identical but interchangeable by commercial custom. The author is grateful to Guy Morton for drawing out this point.

[39] See s I.4.6 of the CMO Reference Manual.

[40] Under the terms of a Dematerialisation Agreement and a Deed of Covenant entered into by all CMO members.

[41] In the deed of covenant, issuers agree that holders of dematerialised CDs shall have the same rights against them as if the CDs had been in physical form.

[42] See the Dematerialisation Agreement.

[43] The Contracts (Rights of Third Parties) Act 1999 permits contractual benefits to be conferred, but not contractual burdens imposed, on third parties.

In some cases, members hold the MMIs beneficially for clients as custodian. Such a custodian member might transfer client MMIs within the CMO without authority. In these circumstances, the client may wish to assert their property claims against the transferred MMIs in the hands of the transferee member or a third party.[44] The contractual provision discussed would not prevent such a claim being brought.[45]

D. Dematerialisation in CREST

A major trend in the international financial markets during recent decades is the consolidation of securities settlement. In 1998 the Bank of England published its *Securities Settlement Priorities Review*. At that time, three settlement systems operated in London, namely CREST (for equities and other corporate registered securities); the Central Gilts Office ('CGO') for gilts;[46] and the CMO. The Review recommended that the three systems be integrated.

The CGO was merged with CREST in 2000. The legal aspects of this merger were relatively straightforward, because of similarities in the legal and operational structures of each of the two original systems.[47] However, fundamental legal and operational differences between the CMO and CREST delayed their merger. Assets settled through CREST are intangible; no physical instruments (or certificates) exist in respect of them. Moreover, all the securities of any one type are fungible. For example, while the accounts of CREST may record the quantity of ICI ordinary shares credited to each member, no note on the accounts of CREST or elsewhere serves to distinguish the ICI ordinary shares of one member from those of another.[48] It is generally accepted that, without special statutory provision, the negotiable status of MMIs would not survive in an intangible and fungible environment, for the reasons discussed in Part II.

The merger is now proposed to take place before the end of 2003. MMIs will be created by deed poll executed by the issuer, and will take the form of fungible and dematerialised registered securities.

[44] Particularly if the wrongly transferring custodian is insolvent, so that their personal claim against it may be worthless.

[45] Under the Dematerialisation Agreement the wrongly transferring custodian would be obliged to indemnify any CMO member transferee against whom such a claim is brought. However, such a claim may be more likely where the custodian is insolvent; its insolvency may make the right of indemnity worthless.

[46] Gilts are UK treasury securities in registered form.

[47] Like CREST, the CGO settled registered securities on a fungible basis.

[48] Indeed, such distinction would not be possible, because CREST assets are undivided; this is explained in Section IIID.

PART II

IV. INTANGIBLES CANNOT BE NEGOTIABLE INSTRUMENTS

This section will argue that, although intangibles can be subject to property rights, they cannot be subject to possession, and therefore cannot be enjoy the status of negotiable instruments.[49]

A. Nature of Intangibles

Personal assets[50] are divided into two classes, namely tangibles and intangibles. Tangible assets are those which can be touched; they are also known as choses in possession,[51] or chattels. Intangible assets or intangibles are those which cannot be touched because they have no physical existence. Essentially, they are rights of legal action. They are also known as choses in action,[52] or claims. An example of an intangible is a debt.

In general, securities are claims. They comprise rights enjoyed by the holder (the most economically significant of which is usually the right to receive payment),[53] which are generally enforced by action against the issuer.[54]

[49] The term 'negotiable instrument' is used here in the traditional sense to mean that class of instrument enjoying privileged status (and in particular, security of transfer) under the Bills of Exchange Act 1882 and/or commercial practice as reflected in case law.

Of course, financial assets (such as MMIs dematerialised in CREST) which enjoy security of transfer otherwise than by their status as negotiable instruments (see the discussion in Section VI below) might be described as 'quasi-negotiable'. See in particular Micheler, n 2.

[50] Assets are divided into real assets and personal assets. Very broadly speaking, real assets are related to land and personal assets are not related to land. Note however that a leasehold interest in land is treated as a form of personal asset.

[51] 'Chose' is a French legal term meaning thing. Tangibles are choses in possession because property rights in relation to them can be enforced without going to court by the self-help remedy of taking the asset into possession, ie by seizing it.

[52] Intangibles are choses in action because they cannot be enforced by taking possession, but only by taking legal action.

[53] Others include, for example, rights to vote and receive financial information.

[54] Or a third party. For example, the issuer's payment obligations may be guaranteed by a third party; in the case of a bill of exchange the holder may have a right to sue a party who has endorsed the bill.

In relation to the treatment of securities as choses in action, see *Colonial Bank v Whinney* (1886) 11 App Cas 426. See also *Borland's Trustee v Steel Brothers & Co Ltd* [1901] 1 Ch 279 'A share is the interest of a shareholder in the company, measured by a sum of money for the purpose of liability in the first place and of interest in the second, but also consisting of a series of mutual covenants entered into by all the shareholders inter se in accordance with s 16 of the Companies Act 1862 [now s 14 of the Companies Act 1985]. The contract contained in the articles of association is one of the original incidents of the share. A share is not a sum of money settled in the way suggested, but is an interest measured by a sum of money, and made up of various rights contained in the contract, including the right to a sum of money of a more or less amount', per Farwell J at 288.

B. Intangibles can be Subject to Property Rights

This section will argue that intangibles can be subject to property rights. Rights arising in private law[55] are divided into personal rights and property rights. They are generally distinguished as follows. Whereas personal rights are enforceable against persons, property rights are enforceable against assets.[56] In legal history, property rights are derived from proprietary actions, or actions in rem.[57] Justinian defines actions in rem (as opposed to actions in personam, from which personal rights have been derived) as follows:

> A plaintiff may sue a defendant who is under an obligation to him, from contract or from wrongdoing. The personal actions lie for these claims…Or else he may sue a defendant *who is not under any kind of obligation to him* but is someone with whom he is in a dispute about a thing. Here real actions lie. [author's italics][58]

For Justinian, obligations are things.[59] On this basis, an obligation (or intangible asset) may be the subject of an action in rem, provided the defendant is someone other than the obligor. Such an action might be brought where an obligation is held on trust. The beneficiary might sue the liquidator of the insolvent trustee to enforce her rights in the trust asset.[60] As against the trustee's liquidator, the beneficiary's action is in rem, and the purpose of the action is to enforce property rights.[61] Of course, as against the obligor, any action in relation to the obligation would necessarily be personal and not proprietary.[62] Obligations can only be subject to property rights in the hands of third parties.[63]

[55] This is as opposed to rights arising in public law, such as constitutional and human rights.
[56] See for example *Bracton on the Laws and Customs of England*, Thorne Transl. (Harvard University Press, Cambridge Mass. 1968), II, 292; *Blackstone*, n 14, I, 16, 20–21.
[57] See for example FW Maitland, *The Forms of Action at Common Law*, AH Chaytor & WJ Whittaker (eds) (Cambridge University Press 1997) (First published 1909), 1, 2: 'Let us remember one of Maine's most striking phrases, 'So great is the ascendancy of the Law of Actions in the infancy of Courts of Justice, that substantive law has at first the look of being gradually secreted in the interstices of procedure'. [Maine, *Early Law and Custom*, p 389.] Assuredly this is true of our real property law, it has been secreted in the interstices of the forms of action.'
[58] *Justinian's Institutes* (Birks and McLeod (transl), London, Duckworth, 1987) 129, J 4.6.2.
[59] 'Some things are corporeal, some incorporeal … Incorporeal things cannot be touched. They consist of legal rights—inheritance, usufruct, obligations however contracted.' Justinian n 58, J.2.2.pr-2. The reification of obligations by Gauis before Justinian is described by Birks and McLeod as 'a brilliant leap': *Justinian*, n 58, 15.
[60] For example, see *Re Stapylton Fletcher Ltd* [1994] 1 WLR 1181.
[61] There has been a considerable jurisprudential debate about whether the rights of the beneficiary under a trust are personal or proprietary. In particular, see Hohfeld, *Fundamental Legal Conceptions* (New Haven: Yale University Press, 1919). As against the trustee, equity acts in personam. However, as against third parties, the beneficiary has proprietary remedies. This is clear in the context of following and tracing actions, as well as in actions against the insolvency official of the trustee.
[62] On this basis, the personal/proprietary distinction serves to categorise actions and not assets. The same asset may be the subject of both personal and property actions.
[63] See Schroeder, n 2.

C. Negotiable Instrument Assimilated to Tangible Assets

As indicated in Section IVA, securities are generally intangibles. However, an exception is made to this general rule in relation to negotiable instruments such as MMIs. Although of course the economic value of the asset is the right to sue for payment, at law negotiable instruments are assimilated to tangible assets.[64]

a. Legal Fiction

Until the late medieval period, the common law restricted the transfer of intangibles, for reasons of public policy. In a feudal society organised around the personal bonds of fealty, it was considered immoral to traffic in claims.[65] However, tangible assets were transferable.[66] In the 12th century the European mercantile community began to draw on Eastern financial techniques learned during the Crusades, including trading in debt. In order to reconcile this practice with the common law restriction, the mercantile community argued that the transferred asset was not a debt, but rather a piece of paper.[67] By a legal convention (referred to hereafter as 'the legal fiction'), the debt is treated as being locked up in the paper. It follows from this that negotiable instruments are categorised as choses in possession and not choses in action. Thus in hiding, the debt is able to do what it could not do in the open, namely move from creditor to creditor.

b. Rules Predicate Physical Asset

Today, the rules relating to negotiable instruments predicate a physical document, which can be delivered, signed on its face and endorsed.[68] Therefore, it

[64] In relation to bills of exchange, see *Hornblower v Proud* (1819) 108 ER 389; *Cumming v Baily* (1830) 130 ER 1320, 1323; *Edwards v Cooper* (1847) 116 ER 386, 388. In relation to bank notes, see *R v Sadi and Morris* (1787) 168 ER.

[65] This was expressed as a common law policy against champerty and maintenance. See Blackstone, n 14, IV, 134: 'maintenance is ... an officious inter-meddling in a suit that no way belongs to one, by maintaining or assisting either party with money or otherwise, to prosecute or defend it ... This is an offence against public justice, as it keeps alive strife and contention, and perverts the remedial process of the law into an engine of oppression' See also 135: 'Champerty ... is a species of maintenance ... being a bargain with a plaintiff or defendant ... to divide the land or other matter sued for between them, if they prevail at law...' See further II, 442: '... for no *chose* in action could be assigned or granted over, because it was thought to be a great encouragement to litigiousness, if a man were allowed to make over to a stranger his right of going to law.'

[66] Subject to restrictions on off-market transfers. Arguably, the general prohibition in certain continental jurisdictions such as Spain on off-exchange securities transactions reflects this old restriction.

[67] Lacanians may be interested to note that this is a form of metonomy.

[68] See the many references to delivery, signing and endorsement in the Bills of Exchange Act 1882. As physical assets, negotiable instruments may be the subjects of possession-based transactions such as bailment and pledge.

is undesirable that intangibles should be categorised as negotiable instruments, because the result of applying the many rules predicating paper to an intangible would be highly uncertain.

D. Possession and Therefore Negotiable Status Unavailable for Intangibles

The class of negotiable instrument is not closed,[69] and the law merchant is capable of recognising new types of negotiable instrument from time to time. It was argued in Section IVB that negotiable status involves property rights and that intangibles may be subject to such rights. It does not, however, follow that intangibles may be negotiable instruments. They may not, because they may not be subject to possession.

Although they are closely related, the legal concepts of property and possession are distinct from each other. Intangibles may be subject to property but not to possession.[70] Possession at common law is a question of fact as well as of law, and involves the physical control of a tangible asset. The concepts of constructive and legal possession permit the relationship between the possessor and the asset to arise symbolically, indirectly or as a legal entitlement. However in all forms of possession predicate an actual or potential relationship of physical control.

Negotiation involves the transfer of possession of an instrument. Possession involves the (direct or indirect) physical control of an asset. Intangibles have no physical existence and therefore cannot be subject to physical control. It follows that they cannot be possessed and therefore cannot be negotiated.[71]

V. FUNGIBLES CANNOT BE NEGOTIABLE INSTRUMENTS

The previous section argued that, although intangibles can be subject to property rights, they cannot be subject to possession, or (therefore) enjoy negotiable status. This section will argue that the position is similar for fungibles. Although they may in certain circumstances be subject to coproprietary rights, fungibles cannot (it will be argued) be individually possessed,[72] and therefore cannot be categorised as negotiable instruments.

Fungibility relates to delivery. Where a loan or custody arrangement is fungible, fungibility relates to redelivery. An asset reveals its fungible or

[69] *Goodwin v Robarts* LR 8 QB 374.
[70] See for example Bracton n 56, I, 121: '… incorporeal things cannot be possessed …'.
[71] Thus it is argued that an intangible cannot be negotiable. However, the author is grateful to Guy Morton for pointing out that an intangible may be an instrument.
[72] Guy Morton correctly points out that banknotes can be individually possessed. Therefore it would be more correct here to say that undivided fungibles cannot be individually possessed.

non-fungible nature when a party (X) loans, bails or otherwise delivers it to a counterparty (Y) subject to a redelivery obligation. The redelivery obligation for a fungible will be equivalent and not in specie.[73]

The following discussion will make a number of distinctions between different types of fungible arrangement. These are relevant to the analysis of property and possession that follows.

The first distinction is whether or not Y is obliged to obtain the redelivered asset from a defined pool. If it is, the arrangement is called 'ex bulk'. If it does not, so that Y is not restricted as to the source of the redelivered asset, the arrangement will be called 'generic'.[74]

A. Ex Bulk

With ex bulk fungibles, Y is subject to a number of obligations. It must (1) maintain a pool of like assets; (2) place the assets it receives from X into the pool; (3) redeliver to X the same number of like assets at the end of the arrangement; and (4) obtain the redelivered assets out of the pool. Commercial examples of such arrangements include granaries[75] and investment securities custody operations.[76]

A pool or bulk may arise in different ways. It may comprise a mixture of different assets of the same kind. Alternatively, it may comprise a single undivided asset, from which an unallocated fraction is redeliverable.[77] (Such a single asset is called a pool or bulk because the claims of Y relate to part of it only.) These alternatives will be considered in turn.

a. Mixtures[78]

Two types of mixtures are distinguished, namely granular and fluid mixtures. In Roman law, these are called, respectively, commixtio and confusio.[79]

[73] See the Justinian distinction between commodatum and mutuum: 'Next, someone who is given something to use—that is to say, who receives a loan of the kind called commodatum—also comes under an obligation by conduct. He is liable to the action of loan for use. His position is quite different from the borrower of a mutuum. Here the thing does not become the property of the borrower, which means he must restore the very thing borrowed.' *Justinian*, n 58, 107.
[74] See *Goldcorp Exchange Ltd (in receivership)* [1994] 2 All ER 806 at 814.
[75] *South Australian Insurance Co Ltd v Randell* (1869) 6 Moo PCCNS 341.
[76] Where a bank or other institution holds securities as custodian for clients, it is customary for the custody agreement to provide that the like assets of different clients may be commingled in the hands of the custodian, so that the redelivery obligation of the custodian to the client is equivalent and not in specie.
[77] The granary provides an example of the first. Fungible custody of registered securities provides an example of the second; this is on the basis that an issue of registered securities is undivided, as argued in Section VCb.
[78] See Peter Birks, 'Mixtures', in Palmer and McKendrick (eds), *Interests in Goods*, 2nd edn, (London, LLP, 1998) 227.
[79] This distinction is not so clearly drawn out in English law: see Blackstone, n 14, II 405.

In a granular mixture, the units comprising the mixture retain their original individual identity. However, they are not held separately, and no record is kept that serves to distinguish one from another. Because each asset is alike, there is no evidence of which is which. Thus the units making up the mixture have individual identity, but it is unknown, as with identical twins whom one cannot tell apart. With granular mixtures, fungibility is epistemologically[80] but not ontologically[81] necessary. Examples include mixtures of grain in a granary, sheep in a flock.[82]

However, in a fluid mixture, the mixing brings the individual identity of the original assets to an end. They merge in the mixture to form a new single asset.[83] The parts of the mixture retain no individual identity. Redelivery out of the mixture is necessarily fungible, and such necessity is ontological. Examples include mixtures of wine in a vessel[84] and oil in a tanker.[85]

b. Undivided Asset

A mixture (whether granular or fluid) occurs when individual assets are combined. The difference between granular and fluid mixtures relates to the composition of the mixture.[86] In either case, more than one original asset predates the mixture. In contrast, a pool or bulk may be formed out of a single asset which has never been divided. As with a fluid mixture, redelivery out of such a pool is necessarily fungible as an ontological matter.[87]

[80] Epistemology concerns how knowledge is established.
[81] Ontology concerns being, or the essential nature of things.
[82] *Justinian*, n 58, 59.
[83] Of course, commentators have pointed out that, at the level of atomic physics, the constituents of certain fluid mixtures (suspensions but not emulsions) remain distinct. The answer to this is that the test is legal and not empirical. This would not be the first time that legal fact differs from empirical fact.
[84] *Justinian*, nn 58, 58.
[85] *Indian Oil Corp ltd v Greenstone Shipping SA, The Ypatianna* [1988] QB 345, [1987] 3 WLR 869, [1987] 3 All ER 893.
[86] Ie, whether it comprises a single asset or multiple assets.
[87] However, see the argument by Professor Sir Roy Goode that a part of an undivided asset cannot be termed fungible, because '… fungibility involves a choice between legally interchangeable units … It therefore presupposes that the asset said to comprise the units or bulk is of a kind that is divisible into separate units capable of being separately owned. … Co-ownership of a single, indivisible asset is therefore incompatible with the concept of fungible units, for the subject-matter of any disposition is a share in the single asset itself by way of co-ownership, and there is no choice to be made.' Roy Goode, n 36, 99. He argues that fungibles must be units which can be differentiated, but which are interchangeable. Because shares are not units, Professor Sir Roy Goode argues, they are not fungible (102). (Incidentally, the equation of intangibles with undivided assets recalls the following passage from Descartes' Sixth Meditation: 'To begin this examination, then, I here remark firstly, that there is a great difference between mind and body, in that body, by its nature, is always divisible and that mind is entirely indivisible. For in truth, when I consider my mind, that is to say myself in so far as I am only a thinking thing, I can distinguish no parts, but conceive myself as one single and complete thing. And although the whole mind seems to be united to the whole body, yet, if a foot, or an arm, or any other part, is separated from my body, it is certain that, on that account, nothing has been taken away from my mind, nor can the faculties of wishing, feeling, perceiving, etc., properly be called its parts, for it is the same mind that is occupied, whole and entire, in willing, perceiving conceiving, etc.

It will be argued below in this section that holdings of registered securities such as shares are an example of such undivided assets.[88]

The two most common types of registered securities are debt and equity securities. Under the terms of issue of registered debt securities, the issuer enters into only one set of covenants in favour of holders.[89] As the covenant creates the debt, there is only one debt for the entire issue, in which holders enjoy fractional interests.[90] Thus, 'There is no identity in stock.'[91] In the case of equity, the rights of shareholders arise under the memorandum and articles of association of the issuer.[92] An issuer has only one set of constitutional documents, and it follows that it is under only one set of obligations to shareholders. This is consistent with the balance sheet treatment,[93] legal analysis,[94] historic development[95] and indeed name of shares,[96] all of which indicate fractional interests in the undivided capital of the issuer.[97]

But it is quite the opposite in corporeal or extended things, for these is none that I cannot easily take to pieces in thought, none that my mind cannot divide very easily into several parts, and, consequently none that I do not know to be divisible. This would suffice to teach me that the mind or soul of man is entirely different from the body, if I were not already convinced of it on other grounds.' *Meditations*, transl. F E Sutcliffe, (Penguin 1968), 164.

The author agrees that, where A holds 950 shares in XCo, then as between A and X none of the 950 shares can be differentiated from another. However, the position differs when third parties become involved. Suppose A transfers half of her shares to B and half to C. It subsequently transpires that 100 of the shares previously held by A were subject to the property claims of D under an equitable tracing action. It is necessary to determine whether D should bring her action against B or C. For this purpose, the rules of equity differentiate between parts of the former holding of A, so that the shares which formed the subject of one of the two transfers are treated as being subject to D's continuing property rights. In effect, they are differentiated by reference to their deemed origin.

This point does not affect Professor Sir Roy Goode's argument, which relates to the certainty of subject matter in the creation of a trust. The author hopes he will forgive her for persisting in using the term 'fungible' in a wider sense, to include parts of undivided assets.

[88] The following discussion assumes that all securities in an issue are pari passu and fully paid. It would be most unusual for securities which are listed and/or actively traded in the secondary market to be otherwise.

[89] Equally, any guarantor will give only one guarantee. Parallel covenant may be given to a trustee for noteholders or bondholders in issues where a trustee is appointed. In these cases, however, enforcement restrictions on the covenants given directly to holders will prevent double payment.

[90] See Goode, n 36, 103.

[91] *Bank of England v Cutler* [1907] 1 KB 889, at 909 per Lawrence J.

[92] As these are given effect by company law.

[93] In general there is no allocation between the two sides of the balance sheet.

[94] See *Ind's Case* (1872) 7 Ch App 485, which held that shares do not differ one from the other, notwithstaning that different numbers are allocated to them. In practice, in any case, shares which are listed and/or traded in the secondary market are unnumbered.

[95] The deed of settlement companies from which incorporated companies are derived were, at law, partnerships with transferable shares. PS Atiyah, *The Rise and Fall of Freedom of Contract* (Clarendon Press, Oxford 1979) 562. Partnership property is held for all partners as joint tenants, and is thus undivided. This accords with the historic term, joint stock company, indicating the interest of investors in the capital of the company is joint, ie undivided.

[96] The natural meaning of the word 'share' is undivided participation. To describe an equity security as a share is not to pun, but to use the word 'share' in its natural meaning.

[97] See also Goode, n 36, 102, 103. Of course, as Guy Morton correctly points out, shareholders do not have property interests in the assets of the company appearing on the other side of the balance sheet.

The securities accounts maintained in the names of CREST members together comprise the register of members of the issuer in relation to the securities credited to them.[98] Title to registered securities (within and outwith CREST) is determined by registration.[99] Accordingly book entry transfer in CREST serves to transfer legal title. Because the legal basis of book entry transfer through CREST is registration, only registered securities may be settled through CREST. Dematerialised MMIs will accordingly take registered form. It was argued in Section VAb that registered securities are undivided assets. On this basis, each type of dematerialised MMIs will form a single, undivided asset, and dematerialised MMIs will be, by ontological necessity, fungible.

Further, dematerialised MMIs will be ex bulk fungibles. As indicated in Section IIID, credits to members' CREST accounts comprise the (only) root of title. In effect, the credit entries constitute the securities. Therefore, the sum of these credit entries together constitutes the pool of MMIs to which the property rights of each relevant member relate.

B. Generic Fungibles

For completeness, this section will discuss generic fungibles. It will be brief because, as indicated in Section VA, dematerialised MMIs are ex bulk. It was seen in Section VA that, with ex bulk fungible arrangements, Y has four separate obligations. In generic arrangement, Y has only one obligation. This is to redeliver the same number of like assets. Y is free to obtain the redelivered assets from any source, and to obtain them just in time to meet the redelivery obligation. During the currency of the arrangement, Y is not obliged to hold assets of the relevant type. This means of course that Y is able to use the assets originally received for its own purposes. This right of use is generally the commercial function of such arrangements.

To the author's knowledge, generic fungible arrangements were originally developed in relation to deposit taking business. The predecessors of contemporary banks were goldsmiths, whose business required them to have strong boxes. The goldsmiths supplemented their income by offering safekeeping services for clients' gold. In due course, they supplemented their income further by lending the gold they held in safe custody to third parties, and charging and retaining interest on those loans.[100] The courts

[98] These arrangements for the electronic transfer of title or ETT were introduced in November 2001 under the Uncertificated Securities Regulations 2001.

[99] In relation to certificated securities, see *Société Generale de Paris v Walker* (1895) 11 App Cas 20.

[100] See the discussion in R Guynn, *Modernising Securities Ownership, Transfer and Pledging Laws* (IBA, 1996).

protected the early development of the lucrative business of banking by upholding this right of use,[101] and its necessary implication of generic fungibility.

C. Ex Bulk Fungibles can be Subject to Property Rights

Property rights predicate assets to which they can relate. It is therefore clear that property rights require their subject assets to exist.[102] The question however arises whether they also require their subject assets to be allocated, so that a particular property right attaches to a particular asset. If such allocation is not necessary, it may be possible for fungible assets to be subject to property rights. This section will consider whether X can successfully claim to enjoy property rights in relation to the delivered assets (or other assets in their place) after delivery to Y and prior to redelivery. Of course, much depends on the drafting of the agreement governing the transaction. If this contains provisions purporting to give X a continuing property interest,[103] can such provisions take effect in accordance with their terms? It will be argued that the answer to this depends on whether or not the fungible assets are ex bulk.

This section will consider in turn mixtures and undivided assets, before turning to charges and the question of certainty of subject matter on the creation of a trust over an unallocated part of a pool. These topics provide support for the proposition that ex bulk fungibles can be subject to property rights. Finally, the structure of such rights is considered at the end of this section.

a. Mixtures

With mixtures, X owns individual assets outright prior to the pooling. The question is whether her property rights survive the pooling, or are extinguished by it. Provided that the mixing is accidental,[104] the effect of mixing

[101] *Carr v Carr* (1811) 1 Mer 625; *Foley v Hill* (1848) 2 HL Cas 28; *South Australia Insurance Co v Randell* (1869) 16 ER 775; *Joachimson v Swiss Bank Corporation* [1921] 3 KB 110; *Space Investments Ltd v Canadian Imperial Bank of Commerce Trust Co (Bahamas) Ltd* [1986] 1 WLR 1072, [1986] 3 All ER 75, [1986] BCLC 485.

[102] Purported assignments of future assets take confer mere contractual rights until the subject assets come into existence: *Holroyd v Marshall* (1862) 10 HL Cas 191.

See also Hobbes *Leviathan*, (London, Penguin, 1985), Part I, ch 3, 97: 'The Present only has a being in Nature; things Past have a being in the Memory only, but things to come have no being at all; the Future being but a fiction of the mind ...'

[103] However, it is increasingly customary to provide that , where a firm exercises a right of use, the client ceases to enjoy property rights in the subject assets, and instead acquires mere contractual rights of equivalent redelivery.

[104] 'The consensual mixing of divided assets may result in co-ownership of the pool if so agreed by the parties; however, this is the result of the agreement and not of the mixing'.

in Roman[105] and in English[106] law is that the owners of the original assets (including X) cease to own individual assets outright, and instead become owners in common of the pool. With mixing, rights of outright ownership are replaced by rights of coownership.

In the case of confusio (but not commixtio), there is also a change in the nature of the assets to which these rights relate.[107] Individual assets are extinguished and a single composite asset is substituted.[108]

b. Undivided Assets

The position differs with undivided assets. The impact of pooling on the property rights of X, and the assets to which they relate, will be considered in turn.

Suppose a custodian, Y, holds a pool of 450 shares for clients. X transfers her 50 shares into the pool, on the basis that she is entitled to receive from Y 50 like shares out of the pool at a later time, and on the further basis that she will enjoy continuing property rights in the meantime.

It was argued in Section VAb that entire issues of shares are undivided assets, so that the rights of the respective shareholders are coproprietary. If, prior to the transfer, X is the registered holder of 50 shares out of an issue of 50,000, she is the coowner (together with all other registered holders) of 1/1000 part of the issue. After the transfer, if Y is the legal holder of 500 registered securities, it is the coowner (together with all other registered holders) of 1/100 part of the issue. After the transfer, X is the coowner (together with all other clients for whom Y holds the pool) of 1/10 part of the coownership rights of Y.[109] For X, first order rights of coownership are replaced by second order rights of coownership. X still has coproprietary rights, but they are at one remove from the underlying asset.

Prior to the transfer, and after it, the asset subject to X's property rights is a single pool or bulk.

On this basis, the proprietary effect of pooling is fairly limited. The nature of the asset does not change, and the nature of the rights does not change (although they are placed at one remove).

[105] *Justinian,* n 58, 2.1.28.
[106] See *Buckley v Gross* (1863) 3 B & S 566, 122 ER 213; *Spence v Union Marine Insurance Co Ltd* (1868) LR 3 CP 427; *Indian Oil Corp Ltd v Greenstone Shipping SA (Panama)*[1988] 1 QB 345*; Glencore International AG v Metro Trading Inc* [2001] 1 Lloyd's Rep 284.
[107] It is important not to confuse property rights with the assets to which they relate. See *Justinian*, n 58 61; Bracton, n 56, II, 48; and Blackstone n 14 Book II, 20, 21.
[108] See *Justinian,* n 58, 2.1.27–28, and the discussion of it by Peter Birks in 'Mixtures', n 78, 233.
[109] The financial markets are familiar with such structures, which rely on the law of trusts. Examples are found among unit trusts, depositary receipts, custodial arrangements and forms of structured finance.

c. Charges

Charges are a form of equitable[110] security interest. It will be argued in this section that they provide indirect support for the proposition that ex bulk fungibles can be subject to property rights. Charges may arise either by the express intention of the chargor, or by operation of law.

Express charges may be fixed or floating. The difference between floating charges and fixed charges was considered in the Court of Appeal by Millett LJ (as he then was) in *Re Cosslett (Contractors) Ltd*:[111]

> The essence of a floating charge is that it is a charge, not on any particular asset, but on a fluctuating body of assets which remain under the management and control of the chargor, and which the chargor has the right to withdraw from the security despite the existence of the charge. The essence of a fixed charge is that the charge is on a particular asset or class of assets which the chargor cannot deal with free from the charge without the consent of the chargee.[112]

A debt of £5,000 may be secured by a fixed charge on an asset worth £100,000. However, fixed charges are not relevant to the discussion of property rights in fungibles. This is because a fixed charge attaches to the whole of the charged asset, even through it may not exhaust it economically. The equity of redemption is a concurrent property interest also attaching to whole of asset.[113]

In contrast, a floating charge does not attach to the charged assets, but hovers over them.[114] Recent case law confirms that a floating charge

[110] In the past, the decisions of different courts generated different bodies of law. The common law courts generated the rules of common law, and the Chancellor's courts generated the rules of equity. (Equally, the mercantile courts developed the rules of the law merchant.) Following a process of consolidation of the different courts (and therefore of the different bodies of law) culminating in the Judicature Acts of 1874–5, English case law is now a unified whole. However, different bodies of rules remain forming different branches of law, so that the rules of common law (which govern legal title to assets) differ from the rules of equity (which govern equitable interests in assets).

At general law, a charge can only take effect in equity. Legal charges over land are a statutory creation under the Law of Property Act 1925 ss 86, 87. See RM Goode, *Legal Problems of Credit and Security* 2nd ed (London, Sweet & Maxwell, 1988) 14–15.

[111] [1998] 2 WLR 131.

[112] At 143.

[113] The equity of redemption is a property right retained by a mortgagor or chargor in the mortgaged or charged asset. It entitles her to the return of the asset upon discharge of the secured obligation.

[114] Romer LJ famously commented (in *Re Yorkshire Woolcombers Association Ltd* [1903] 2 Ch 284 at 295):

> …I certainly think that if a charge has the three characteristics that I am about to mention it is a floating charge. (1) If it is a charge on a class of assets of a company present and future; (2) if that class is one which, in the ordinary course of the business of the company, would be changing from time to time; and (3) if you find that by the charge it is contemplated that, until some future step is taken by or on behalf of those interested

confers a property right on the chargee.[115] This supports the proposition that property rights need not attach to particular assets, and therefore that ex bulk fungibles may be subject to property rights.

In certain circumstances, a charge arises by operation of law. These include the earmarking of funds for the discharge of a debt, and (in some circumstances) equitable tracing through a pool.

In *Swiss Bank Corporation v Lloyds Bank Ltd*[116] it was held that the earmarking of a fund for the discharge of an obligation may serve to confer an equitable property right in the fund in the nature of a charge. This also indirectly supports the possibility of property rights in ex bulk fungibles (the bulk in this case being the earmarked fund).

Depending on the circumstances, tracing may take place either under the rules of common law or the rules of equity.[117] Equitable tracing is a process[118] which becomes available to beneficiaries when their assets have been disposed of by a trustee or other fiduciary in breach of duty. It enables the beneficiaries to identify the assets which are subject to their property claims in the hands of third parties (either in their original form or in the

> in the charge, the company may carry on its business in the ordinary way so far as concerns the particular class of assets I am dealing with.

On appeal in the same case, Lord Macnaghten commented (in *Illingworth v Houldsworth* [1904] AC 355 at 358):

> I should have thought that there was not much difficulty in defining what a floating charge is in contrast to what is called a specific charge. A specific charge, I think, is one that without more fastens on ascertained and definite property or property capable of being ascertained and defined; a floating charge, on the other hand, is ambulatory and shifting in its nature, hovering over and so to speak floating with the property which it is intended to affect until some event occurs or some act is done which causes it to settle and fasten on the subject of the charge within its reach and grasp.

Neither floating charges nor property rights in relation to ex bulk fungibles attach, but (it is argued) the want of attachment differs as follows. The defining characteristic of a floating charge (as opposed to a fixed charge) is the freedom of the chargor to remove assets from the pool without the consent of the chargee. Thus the composition of the pool may change from time to time, so that the pool subject to the charge does not attach to particular assets. At any time, the charge relates to all the assets in the pool (although it may not economically exhaust them, ie, the value of the pool may exceed the value of the secured claim).

In contrast, the rights of X in relation to ex bulk fungibles relate only to an unallocated part of the pool. Therefore, the want of attachment for ex bulk fungibles is internal to the pool; whereas the want of attachment for floating charges is coterminous with the pool.

[115] Again, by Millett LJ in *Cosslett*. For the earlier position, see the discussion in *Swiss Bank Corporation v Lloyds Bank Ltd* [1979] 1 Ch 548; [1982] AC 584.

[116] [1979] 1 Ch 548; [1982] AC 584.

[117] Because the asset of a trust beneficiary is equitable, she must bring any tracing action for it in equity. Where the legal owner of an asset is wrongfully deprived of the asset, she may be able to bring a tracing action at law. See *Trustees of the Property of F C Jones & Sons v Jones* [1996] 3 WLR 703.

[118] 'Tracing properly so-called, however, is neither a claim nor a remedy but a process.' *Boscawen and Others v Bajwa and Another* [1996] 1 WLR 328, per Millett LJ at 334. See also *Foskett v McKeown* [1997] 3 All ER 392, CA.

form of substitute assets).[119] The key point for the present discussion is that tracing in equity (unlike tracing at common law) can follow assets through a pool in which they are mixed with other assets.[120] Detailed rules have been developed in case law to identify trust assets when they have been mixed with the assets of the trustee[121] and the assets of beneficiaries or innocent third parties.[122]

Property rights associated with an equitable tracing claim can survive the loss of attachment to particular assets that occurs upon pooling. This (indirectly) supports the proposition that ex bulk fungibles can also be subject to property rights.

d. Certainty of Subject Matter

Further indirect support for the proposition that ex bulk fungibles may be subject to property rights is provided by the case law relating to certainty of subject matter upon the creation of a trust. Three matters must be certain[123] in order for a valid trust to arise.[124] These are certainty of intention, certainty of beneficiary and certainty of subject matter. A number of cases relate to certainty of subject matter, and deal (in effect) with the question of whether a valid trust can be created over fungibles, or whether such a purported trust fails for want of certainty of subject matter.[125] (The issue is similar to that arising in cases concerning mixing, but differs as to timing.[126])

It is clear that where the fungibles are ex bulk, the subject matter of the trust is sufficiently certain.[127] (It will be seen that the position differs for

[119] *Lipkin Gorman v Karpnale* [1991] 2 AC 548, at 573 per Lord Goff.

[120] The general rule is that common law tracing cannot: '… at common law, property in money, like other fungibles, is lost when it is mixed with other money.' *Lipkin Gorman*, per Lord Goff at 572. See also *Re Diplock* [1948] Ch 465 at 520.

[121] See in particular *Re Hallett's Estate* (1880) 13 Ch D 695, *Re Oatway* [1903] 2 Ch 356, *Roscoe v Winder* [1915] 1 Ch 62.

[122] See in particular *Clayton's Case* (1816) 1 Mer 572), *Re Diplock* [1948] Ch 465, *Barlow Clowes International v Vaughan* [1992] 4 All ER 22 and *Russell-Cooke Trust Co v Prentice & ors* (2002).

[123] (or ascertainable).

[124] *Knight v Knight* (1840) Beav 148, 9 LJ Ch 354.

[125] The same issue arises is law as well as in equity. In relation to the sale of goods, the old rule that property does not pass prior to ascertainment was modified by the Sale of Goods (Amendment) Act 1995, which (very broadly) provides that title to goods in ex bulk sales passes on payment of the purchase price.

[126] The rules relating to mixing discussed in Part VAa do not answer the allocation question, because they apply in cases where the individual ownership of different assets predates the pooling of those assets. In cases concerning certainty of subject matter, the timing is reversed because the pool predates purported property rights in it. See *inter alia* Benjamin, *The Law of Global Custody* (London, Butterworths, 1996), 45 and Peter Birks, 'Mixtures', n 78, 228.

[127] See *inter alia Stapylton Fletcher (in administrative receivership), Re Ellis Son & Vidler (in administrative receivership)* [1995] 1 All ER 192; *Hunter v Moss* [1993] 1 WLR 934; [1994] 1 WLR 452; *Harvard Securities (in liquidation), Holland v Newbury and another* [1997] 2 BCLC 369; *Re CA Pacific Finance Ltd (in liquidation) and another* [2000] BCLC 494. *Hunter v Moss* related to ex bulk fungibles, unlike Goldcorp which inter alia considered the question of generic fungibles.

generic fungibles.) As discussed in the following section, the precise structure of the property rights of beneficiaries is still open to debate.

e. Structure of Ex Bulk Property Rights

It is clear from the preceding discussion that property rights can exist in relation to ex bulk fungibles. This section will consider their form. The options available include (and may be limited to) coownership and charge.

A similarity between charges and coownership rights is that they may relate to but not exhaust a pool. A charge is an encumbrance, which (like coproprietary rights) forms part of, but does not exhaust, the property rights in the subject assets. The rights of the chargee are concurrent with the rights of the chargor under the equity of redemption. However, the rights conferred by a floating charge are not the same as coproprietary rights. There is not necessarily more than one floating chargee). Both the charge and the equity of redemption relate to the entire pool. A further and crucial difference between a charge and coownership relates to changes in the value of the pool. Coowners but not chargees take the profits associated with increases, and suffer the losses associated with reductions, in the value of the subject assets.

It was argued in Section VA that mixtures and fungible right in relation to undivided assets give rise to coproprietary rights. The property rights arising in relation to earmarked funds[128] and pursuant to equitable tracing may take the form either of a charge or coproprietary rights (and in the latter case, the claimant may elect which form her rights are to take).[129]

There is authority that a beneficiary's interests in an unallocated part of a pool takes effect as a coproprietary right over the whole pool.[130]

[128]See *SBC v Lloyds* [1979] 1 Ch 548; [1982] AC 584.

[129]Where the traceable trust fund has been mixed with other assets (for example where the proceeds of sale of the original trust assets form part of the purchase price of a new holding of securities), the plaintiff may claim an equitable charge on the amalgam, or a proportionate share of it: *Re Tilly* [1967] 2 All ER 303, at 310 per Ungoed-Thomas J. The former option would be more valuable to the plaintiff where the new holding has dropped in value; the latter option would be more valuable where it has appreciated.

[130]All beneficiaries enjoy their rights under this single trust, and together beneficially co-own the pool. There is no uncertainty of subject matter, because the trust property comprises the entire pool. See *Re Stapylton Fletcher Ltd (in administrative receivership), Re Ellis Son & Vidler Ltd (in administrative receivership)* [1995] 1 All ER 192, at 210 per Judge Paul Baker QC. *Stapylton Fletcher* related only to legal interests arising in the sale of goods; *Re London Wine (Shippers) Ltd* [986] PCC 121 considered the position both at law and in equity, and indicated (per Oliver J at 137), that very clear wording would be required in order for an equity in common to arise.

An attraction of this approach is that the legal results of consensual mixing and the allocation question coincide; coownership rights arise in either case. Moreover, in a wide range of repackaging arrangements (including depositary receipts) the interests of investors arise under express co-ownership provisions. The legal analysis of the securities markets is tremendously

There are judicial indications that this is not the position;[131] unfortunately, these judgments do not suggest an alternative structure.

f. Generic Fungibles

However, if fungibles are generic, they may in general not be subject to property rights. Continuing property rights require the continued existence of assets to which they can relate.[132] If no fund of assets exists, property rights cannot arise.[133] If the fund's existence is interrupted, property rights are extinguished.[134] (This may be the historic jurisprudential basis for the principle that cash at bank confers personal and not property rights on the depositor.)

D. Negotiable Status Unavailable for Fungibles

It was seen that property rights can extend to ex bulk fungibles through the concepts of coownership and charge. The law of possession is less developed than the law of property,[135] and there is old authority that a fungible cannot be possessed.[136] There is no possessory concept equivalent to the charge.

simplified if the legal nature of interests in securities arising under such repackagings, and those arising under custodial and settlement arrangements, are understood to be the same.

The co-ownership approach raises what Professor Birks has called the problem of unilateral partition. If client assets are co-owned, does each client require the consent of the other clients to dispose of her interest? Birks, n 78, 229.

In relation to goods, a statutory solution is provided in the new sections 20A and 20B of the Sale of Goods Act 1979. These provisions were introduced by the Sale of Goods (Amendment) Act 1995, which provides (in 20A) for implied co-ownership and (in 20B) for unilateral partition. Of course, intangibles are not goods, and they fall outside the scope of these provisions. However, Professor Birks suggests that '… the day may not be far off when a solution on those lines …' will be judicially developed to provide a general solution to the problem of unilateral partition.

[131] See *Hunter v Moss* [1993] 1 WLR 934; [1994] 1 WLR 452; *Harvard Securities (in liquidation), Holland v Newbury and another* [1997] 2 BCLC 369; and *CA Pacific Securities Ltd (in Liquidation)* 2000 BCLC.

[132] 'When something has ceased to exist it is no longer possible to bring a vindication …' *Justinian*, n 58, 57. This principle informs cases relating to tracing in general (*Re Hallett's Estate, Knatchbull v Hallett*, (1880) 13 Ch D 696), tracing through an overdrawn account (*Bishopsgate Investment Management Ltd v Homan* [1995] Ch 211; *WestdeutscheLandesbank Girocentrale v Islington London Borough Council* [1996] AC 669) and the lowest intermediate balance principle in tracing (*Roscoe v Winder* [1915] 1 Ch 62 and *Pinkett v Wright* (1842) 67 ER 50).

[133] *Re London Wine (Shippers) Ltd* [1986] PCC 121; *Re Goldcorp Exchange Ltd (in receivership)* [1995] 1 AC 74. See also the rejection of the concept of a general charge proposed in *Space Investments Ltd v Canadian Imperial Bank of Commerce Trust Co (Bahamas) Ltd* [1986] 3 All ER 75, in *Re Goldcorp Exchange Ltd (in receivership)* [1995] 1 AC 74.

[134] In tracing, or upon the creation of fluid mixtures, the concept of substitution permits property rights to continue despite a change in the asset to which they relate. However, any interruption in the existence of the subject assets serves to extinguish the property right.

[135] The law of possession is fully expounded in Pollock, n 24. Arguably this remains the leading work. See also *Interests in Goods*, n 75.

[136] *South Australian Insurance Co Ltd v Randell* (1869) 6 Moo PCCNS 341; *USA v Dollfus Mieg et Compagnie SA* [1952] AC 582, HL.

However, copossession appears to be a possibility, and more recent case law suggests that possession may extend to ex bulk fungibles.[137]

Nevertheless, even if a fungible may be owned and possessed, there is no authority a fungible asset can be a negotiable instrument. It is considered conceptually impossible that it should. This is because the existing law relating to negotiable instruments refers extensively to signature, delivery and endorsement, each of which predicate an allocated asset.

PART III

VI. SECURITY OF TRANSFER

It was indicated in Section IIA that the chief advantage of negotiability for the holder of an instrument is security of transfer, and further (in Sections IV and V) that a dematerialised MMI cannot be a negotiable instrument. It follows that the impact of dematerialisation on the security of transfer of MMIs must be assessed.

A. Nature of Security of Transfer

A transfer is secure if the good faith purchaser takes the transferred asset free from adverse claims. Adverse claims may come from three sources, namely the liquidator of an insolvent transferee, the issuer of the transferred asset, and third parties.

If the transferor is affected by insolvency at the time of the transfer or shortly thereafter, the transfer may be affected by mandatory provisions of insolvency law designed to protect general creditors of the insolvent.[138] Transfers made after the commencement of insolvency proceedings are generally void.[139] In addition, transfers made within six months of insolvency[140] may be set aside by the liquidator of the transferor if they amount to transactions at an undervalue[141] or preferences.[142] Such avoidance is

[137] *Harding v IRC* [1976] 1 NZLR 337; *Mercer v Craven* unreported 12 February 1993, CA affd [1994] CLC 328, HL.
[138] 'Fairness in liquidation demands that the general body of creditors be protected from dispositions of the company's assets in the period leading up to liquidation which confer improper advantages on certain creditors or other parties. It demands that the collective nature of the insolvency process be protected. The law on the avoidance of transactions, accordingly, seeks to protect collectivity and the principle of *pari passu* distribution and to deal with the unjust enrichment of a particular party at the expense of the general body of creditors.' V Finch, *Corporate Insolvency Law, Perspectives and Principles*, (Cambridge University Press 2002), p 398.
[139] Insolvency Act 1986 s 127, relating to post insolvency dispositions.
[140] Or two years if the parties are connected: Insolvency Act 1986 s 240.
[141] Insolvency Act 1986 s 238.
[142] Insolvency Act 1986 s 239.

known as insolvency displacement. Transfers of negotiable instruments have never enjoyed special protection from insolvency displacement. However, in the wholesale financial markets, insolvency displacement may have a systemic impact.[143] Therefore a special policy case is made for protecting both transactions in financial assets, and deliveries made pursuant to such transactions. Such protections are available under a number of statutory provisions.[144] The dematerialisation of MMIs will not affect the availability of these protections.

In contrast, the legal position of the transferee in relation both to issuer's claims and third party claims will be affected by the loss of negotiable status with the dematerialisation of MMIs.

B. Issuers' Claims[145]

a. Background

This section will consider the background to issuers' claims; the succeeding sections will consider how they may be addressed.

When a MMI is issued, a contract is formed between the issuer and the first holder ('A'), which inter alia obliges the issuer to make payment in accordance with the instrument's terms. When the MMI is transferred from A to B, B is concerned to ensure that the obligation of the issuer to pay is not affected by any circumstances relating to the original contract between the issuer and A. For example, the issuer's payment obligation under the MMI should not be subject to set off against any debt owed by A to the issuer, or rendered subject to rescission by any misrepresentation made by A to the issuer in relation to the original contract.[146] Such rights of set off or rescission are referred to in this context as 'equities'.

The position of B depends inter alia on the legal nature of the transfer. The term 'transfer' is not a legal term. It refers to a transaction, the economic result of which is that one person ('the transferee') holds the same asset, or an asset equivalent to the asset, previously held by another person ('the transferor'). There are a number of different legal techniques of transfer. Assignment, negotiation and novation are relevant to this discussion.

[143] Also known as the domino effect, whereby one failure causes the failure of other transactions and, ultimately, firms.
[144] See Part VII of the Companies Act 1989; the Settlement Finality Directive (98/26/EC); the Financial Collateral Directive (2002/47/EC); the Winding Up Directive (2001/24/EC); the Insurance Winding Up Directive (2001/17/EC) and the Insolvency Regulation (Regulation 1346/2000/EC).
[145] See generally E Micheler, n 2.
[146] 'Misrepresentation (whether fraudulent, negligent or innocent) gives the representee the option to avoid the contract.' P Birks (ed), *English Private Law* (Oxford University Press, 2000), para 8.174.

Assignment is a technique for transferring contractual rights (but not contractual obligations).[147] The effect of assignment is that the original contract remains in place, but that the assignee takes over the rights of the assignor under it. An assignment can take place at law or in equity.[148] However, any assignment (whether legal or equitable) is 'subject to the equities'.[149] This means that if the transfer from A to B takes the form of an assignment, then under general contractual principles the issuer would be relieved from its obligation to pay B to the extent that any equity relieves it of its obligation to pay A. This does not offer a level of security of transfer acceptable to transferees in the MMI markets.

An alternative technique of transfer, which offers greater security, is negotiation. Negotiation is a modified form of assignment, which is only available in relation to negotiable instruments. Unlike ordinary assignment, negotiation is free from equities. This means that where the transfer from A to B takes effect as a negotiation, B's claims against the issuer are not affected by any equities arising between the issuer and A.

It was argued in Sections IV and V that dematerialised MMIs cannot be negotiable instruments. It is therefore important to determine whether the protection from equities is lost with the loss of negotiable status. If it does, the dematerialisation of MMIs will prima facie reduce security of transfer.

b. Estoppel

It has been argued[150] that the transfer of negotiable instruments free from issuers' equities does not rely on negotiation, but rather on the doctrine of estoppel. Estoppel is a rule of evidence that rests on representation. Very broadly, where a person makes a representation upon which another person relies so as to change their position to their detriment, then (where estoppel operates) the representor is not permitted in litigation to bring evidence of matters contradicting the representation. The MMI bears on its face the representation of the issuer that it will pay the bearer. B purchases the MMI from A in reliance on that representation. Therefore, it is argued, the doctrine of estoppel prevents the issuer from asserting that, notwithstanding its representation, it is relieved in whole or part from its obligation by issuers' equities.

This argument supports the proposition that bearer securities which are not negotiable instruments may nevertheless be transferred free from issuer's equities. However, it cannot assist dematerialised MMIs. This is because the argument rests on the representation of the issuer made on the face of the instrument, and of course a dematerialised instrument has no face.

[147]See *Chitty on Contracts* 28th edn (London, Sweet & Maxwell, 1999), paras 20-075—20-078.
[148]Legal assignments of choses in action are subject to formalities under section 136 of the Law of Property Act 1925.
[149]S 136 and *Business Computers Ltd v Anglo African Ltd* [1977] 1 WLR 578.
[150]Ewart, n 2, Blair, n 2, Micheler, n 2, Benjamin, n 126.

Of course, MMIs which are dematerialised within CREST will be registered securities. Estoppel also operates in relation to certificated registered securities as follows. The issue of a share certificate by a company operates as a representation that the person in whose name the certificate is issued is the shareholder. A transferee from the certificated holder may rely on this representation, and the issuer is estopped from challenging the validity of the transfer on the basis of defects in the transferee's title.[151] It is unclear whether the estoppel can extend to uncertificated securities on the basis that registration amounts to a representation of title.[152] Thus, the doctrine of estoppel may, but does not necessarily, address the problem of issuers' equities in relation to MMIs. (A legislative solution within the Uncertificated Securities Regulations is discussed in Section VID)

c. Novation

An alternative argument is as follows. The problem of issuers' equities is associated with assignment. Assignment serves to transfer rights under a contract from transferor to transferee. Because the rights of the transferee arise under the original contract, they are subject to equities affecting the original contract. However, it is argued that book entry transfers of securities do not proceed by assignment, but rather by novation.

Novation is an alternative technique to assignment for transferring contractual rights; novation also serves to transfer obligations.[153] The term means 'making anew':

> It should, however, be noted that the effect of a novation is not to assign or transfer a right or liability, but rather to extinguish the original contract and replace it by another.[154]

Suppose there is a contract between N and O. O wishes to transfer her interest under the contract to P. This can be achieved by novation if N, O and P agree together that a new contract between N and P will replace the old contract between N and O. The old and new contracts are on the same terms, although with different parties.[155]

The transfer of registered securities operates under a statutory regime.[156] However, it has been argued that the jurisprudential basis for the transfers of

[151] See *Re Bahia and San Francisco Ry Co* (1865) Law Rep 3 QB 118.
[152] See Micheler, n 2, p 374.
[153] As indicated in Part VIBa, assignment is effective only in relation to rights, and not obligations.
[154] *Chitty on Contracts*, n 147, para 20-086, 1076.
[155] They also have different dates, with implications for the rules of priorities and insolvency law.
[156] The primary legislative basis for transfers in CREST is the Companies Act 1989. The statutory regime for transfers of shares outside CREST is the Companies Act 1985.

registered securities is novation.[157] The security comprises a contract between the issuer and the existing holder. By registering a transfer, the issuer implicitly agrees with the existing holder and the transferee to extinguish the old security and in its place to create a new one in favour of the transferee.

The proposition that registration in CREST takes effect by novation is lent further support by the case of *R v Preddy*.[158] It was explained in Section IIID that in CREST, registration takes effect by book entry transfer. *Preddy* concerned the effect of book entry transfers, not across securities accounts but across cash deposit accounts. However, it is relevant to CREST transfers by analogy. The case held that, following a book entry transfer, the asset of the transferee is legally distinct from the asset of the transferor. This is consistent with novation, but inconsistent with assignment.[159]

Because the asset of the transferee differs from the asset of the transferor or other predecessor in title, it is not subject to any equities affecting such prior asset.

Therefore novation proceeds free from issuers' equities.

C. Third Party Claims

The preceding discussion has dealt with insolvency displacement and issuers' claims. Third parties comprise a third source of potential adverse claims facing the good faith purchaser.

a. Security of Transfer v Security of Title

As indicated in Section VIA, the principle of security of transfer provides that a transfer of assets to a good faith purchaser should not be disturbed by adverse claims. This principle reflects the commercial values of a market economy. It potentially conflicts with the principle of security of title, which provides that fraud is not effective to deprive a person of her assets.[160] The conflict between the two principles becomes real in the case of third party claims, as follows.

The cynical view that commercial lawyers take of brokers and other intermediaries to whom client securities are entrusted may be explained by the frequency with which the following scenario recurs in case law. L entrusts her securities to M to hold on her behalf. In breach of duty, M sells (or mortgages) L's securities to N, a good faith purchaser (or mortgagee)

[157] See Benjamin, *Interests in Securities* (Oxford University Press, 2000), 3.06 and Kam Fan Sin, *The Legal Nature of the Unit Trust* (Oxford, Clarendon Press, 1997), 319. See also the detailed discussion in Micheler, n 2, p 363.

[158] [1996] AC 815.

[159] See JE Penner, *The Idea of Property in Law* (Oxford: Clarendon Press, 1997), 147.

[160] This principle supports the stable property relationships on which pre-modern social relations were predicated.

without notice of the breach of duty, and M absconds (or becomes insolvent) with the proceeds (or loan). L discovers the fraud, and sues N for the return of the securities. N naturally wishes to retain them, because she has no prospect of recovering the purchase price (or loan) from the absent (or insolvent) M. In the ensuing dispute between L and N, the court must select a victim from between two innocent parties. The principle of security of transfer favours N, and that of security of title favours L.

The appropriate balance between the two principles is a policy issue that different branches of English law have resolved in different ways.

b. The Different Branches of Law

Depending upon the circumstances of the dispute, the respective rules of common law, the law merchant and equity may apply. Each of these resolves the conflict between security of transfer and of title in a different way.

The general rule is that of common law, which applies unless either the rule of the law merchant or that of equity displaces it. It is that *nemo dat quod non habet*, or no-one can give that which she does not have. Thus, N does not acquire title, because it is not M's to give, and the assets are restored to L.[161]

If the disputed asset is a negotiable instrument, the rule of common law is displaced by the rule of the law merchant. As indicated earlier, this favours security of transfer over security of title. The holder in due course of a negotiable instrument (very broadly, a good faith purchaser)[162] takes free from third party claims. Thus, N takes the asset free from the claims of L.

However, if L is suing in equity, the rules of equity will apply. L may sue in equity if her interest in the disputed asset prior to the transfer was equitable, not legal, for example where her interest arises under a trust. Equity permits L to follow[163] the asset into the hands of N, unless a defence is available. An important defence is that N satisfies the legal definition of 'equity's darling'.[164] Equity's darling is the bona fide purchaser for value of the legal estate, without notice of the breach of duty.[165]

D. Impact of Dematerialisation on Security of Transfer

It was argued in sections 4 and 5 that dematerialisation will remove negotiable status. It follows that the law merchant will not govern third party claims.

[161] A number of exceptions are available to this general rule, but they are not relevant to transfers of dematerialised MMIs.
[162] This term is defined in s 29(1) of the Bills of Exchange Act 1882.
[163] In *The Law of Tracing* (Clarendon, Oxford, 1997) Lionel Smith distinguishes between following original assets in the hands of third parties, tracing substitute assets in the hands of third parties, and claiming such assets: see ch 1 Section II.
[164] Other possible defences are dissipation (see *Re Diplock* [1948] Ch 466) and change of position (see *Lipkin Gorman v Karpnale Ltd* [1991] 2 AC 548).
[165] See *Pilcher v Rawlins* (1872) 7 Ch App 259.

In general, therefore, the position will depend upon whether the interest of the claimant in the asset of which she has been deprived was legal or equitable. If the claimant was a member of CREST,[166] so that her interest was legal, then under the general principles of common law she will succeed. This increases exposure to third party claims, and thus reduces security of transfer.

If, however, the claimant was the custody client of a CREST member whose interest was equitable and not legal,[167] then under general equitable principles her position may depend upon whether the good faith purchaser acquired legal title to the disputed assets, or merely an equitable interest in them. Where the good faith purchaser is a member of CREST acquiring a legal interest, she is a candidate for the defence of equity's darling, and therefore for victory. If, however, she is custody client whose interest is equitable only, she is not.

On the above basis, the general principles of English law do not deal adequately with the question of third party claims. Security of transfer is reduced, and legal risk is increased, as the outcome of any dispute may depend arbitrarily on the operational detail of the parties' custody arrangements. More generally, this reflects the need for a unified law of priorities which does not differentiate between legal and equitable interests.[168] In the specific context of the dematerialisation of MMIs, it suggests the need for a legislative solution.

It has been argued that such a solution is available in the form of regulation 33[169] of the Uncertificated Securities Regulations 2001, and that this serves to prevent adverse claims as effectively as negotiability.[170]

In short, there are strong arguments that the dematerialisation of MMIs does not reduce the security of their transfer.

PART IV

Electronic settlement began as long ago as the late 1960s,[171] and the process of its legal assimilation is well under way. There have been a number of discussions by legal writers.[172] The Bank of England reports that the

[166] Members of CREST having legal title to securities credited to their CREST accounts include sponsored members, ie custody clients who do not themselves have electronic links with CREST, for whom the sponsoring custodian maintains the electronic link and operates the account on their behalf.

[167] This will include all custody clients for whom a custodian holds assets through CREST in the custodian's name, whether their interests are pooled or segregated by the custodian.

[168] See arguments to this effect in J Ulph, 'Good Faith and Due Diligence', in *Interests in Goods*, n 75, and also Law Commission Consultation Paper No 164.

[169] Formerly regulation 29 of the Uncertificated Securities Regulations 1995.

[170] This view was supported by the opinion of Richard Sykes QC. See the discussion in Benjamin, n 157, 213.

[171] Euroclear was established in 1969.

[172] See n 2.

banking community is relaxed about the loss of negotiable status for MMIs.[173] Strong arguments were set out in Section VI that the loss of negotiable status does not reduce security of transfer or otherwise disturb the rights of interested parties.

All of the above might suggest that the dematerialisation of MMIs should be fairly uneventful. However, it is strange to report that there remains tremendous conceptual resistance among some members of the legal community. This is not so much to the legal consequences of dematerialisation as to its operational fact. Anecdotally, Professor Jim Rogers reports how, in the discussions surrounding the revision of Article 8 of the UCC, colleagues became incensed to the point of incoherence when they were told that investors did not own individual securities. In practice, the author and her colleagues have regularly experienced the same thing. Many lawyers in London and New York are distressed to learn that the assets are intangible and unallocated. Even if one offers reassurance in relation to particular legal or operational concerns, there remains a general unease.

The explanation for this may lie in the cultural echoes surrounding dematerialisation. It was seen that, in the pre-modern era, the paper represented the debt; that in the modern era, it constituted it; and that now it is absent. In this part it is argued that this process whereby debt and paper are in turn associated and disassociated has troubling semiotic, philosophical, psychoanalytic and religious resonances.

VII. SEMIOTIC ASSOCIATIONS[174]

Semiotics is the theory of signs.[175] In his *Course in General Linguistics*,[176] Saussure considered language as a system of signs.[177] A word is a signifier,

[173] Consultation among bankers and commercial practitioners on behalf of the Law Commission in 1999 and 2000 by Hugh Pigott and Joanna Benjamin for the Centre for Law Reform indicated that there was little appetite for a new Bills of Exchange Act in relation to electronic trade finance instruments.

[174] See generally Marc Shell, *Money, Language, and Thought*, (The John Hopkins University Press, Baltimore and London 1982). In a series of literary critical essays, Shell demonstrates the likeness between monetary and linguistic structures. He considers in particular the referential role of money, and the increased distance between signified and signifier with the transition from coin to paper money. Shell identifies a comparable slippage between words and their meanings with the transition to the modern era.

[175] In *Foundations of the Theory of Signs* (1938), CW Morris divided semiotics into three branches, namely: syntactics (the study of the relations between signs); semantics (the study of the relations between signs and the things for which they stand); and pragmatics (the study of the relations between signs and those using them). On this basis, the conceptual problems of dematerialisation are semantic. See also *Signification and Significance*, (Cambridge, Mass, 1968).

[176] Ferdinand de Saussure, transl. Roy Harris (Duckworth 1983) originally published from students' lecture notes in 1915.

[177] Signs are divided into three categories by Peirce. These are: icons (in which the signifier visually or otherwise resembles the signified such as a map or portrait); natural signs (where in the normal course of experience the signifier *indicates the presence* of the signified, as spots indicate viral infection and clouds indicate rain); and conventional signs (where interpretation

its meaning is the signified,[178] and the link between them is arbitrary (or conventional).[179] Of course, the link between a credit entry in an electronic securities account and the debt of the securities' issuer is also arbitrary. However, this cannot explain conceptual resistance to dematerialisation, because the link between the debt and a paper instrument makes no greater natural or rational sense: it is equally conventional.[180]

It was seen in Section III that three stages took place in the history of MMIs, as follows: (i) the paper originally signified the debt; (ii) thereafter, the legal fiction of this semiotic relationship was collapsed, so that the signifier was conflated with the signified, ie the paper was treated as being the debt; (iii) with dematerialisation, the paper is removed. It was also seen that the transition from (ii) to (iii) encounters conceptual resistance, and that this resistance is not explained by any increase in semiotic arbitrariness. The first transition from (i) to (ii) involved forgetting that a sign was a sign. The conceptual resistance to the second transition may be explained by the discomfort of being reminded that a sign is merely a sign. (The discomfort is the implication that, even though we have taken hold of the signifier, the signified still eludes us.)

A similar conceptual struggle takes place throughout financial law. It was argued[181] that the subject matter of financial law is merely notional; debts and companies are legal ideas having no non-legal counterpart.[182] The courts can discover nothing about debts and companies empirically.[183] The natural epistemological alternative to empiricism is rationalism, but this is alien to common law.[184] Common law legal reasoning solves this problem by the widespread use of metaphor.[185] However, it takes tremendous intellectual energy always to remember that the relationship

must be learned by reference to a *convention*, such as traffic lights, and non-onomatopoeic words). The paper instruments constituting traditional MMIs neither resemble the debt of the issuer not indicate its presence. On this basis they are conventional signs.

[178] John Locke in his *Essay Concerning Human Understanding* (Clarendon Press, Oxford, 1975) argued that words are signs for the ideas in the mind of the speaker.

[179] There is no natural or rational link between signifier and signified. The link is conventional, and must be learned. 'The thesis of the arbitrariness of the sign is central to Saussurian linguistics ...': P Goodrich, *Legal Discourse* (Macmillan, London, 1987), p 24.

[180] Of course, the economic value of an MMI derives from the right to be paid by the issuer. A banker brings herself no nearer to being paid by clutching a paper instrument than by identifying her name on an electronic account.

[181] Above.

[182] Of course, the relationship between legal ideas and their non-legal counterpart is not straightforward. See the interesting argument that a legal natural person is not, in truth, natural, in Lacey, n 14, 28.

[183] Ie by sensory perception. Of course, we have empirical studies of corporate and credit behaviour.

[184] See the discussion in Section VIII below.

[185] The image of a physical tie etymologically implicit in the word 'obligation' is drawn out in *Justinian,* n 58, J.3.13.pr. See also the discussion of metaphor in relation to corporate personality in Lacey, n 14, 25. Interestingly, Kahn, n 14, p 62 describes money as a form of analogy.

between the legal idea and its metaphorical representation is semiotic and not actual.[186] Some judgments seem to forget this and, in the grip of the metaphor, reach decisions which are arguably harsh to claimants[187] or commercially unhelpful,[188] as if by empirical necessity. In contrast, other judgments controversially transcend the imagery of financial law.[189]

VIII. PHILOSOPHICAL ASSOCIATIONS[190]

Western philosophy is broadly divided into two traditions,[191] namely rationalism (following Plato) and empiricism (following Aristotle). A key difference[192] between them concerns epistemology.[193] More particularly, it concerns the relative epistemological priority of abstract ideas and sensory data. Rationalist thought proceeds *a priori*. This means that it begins by assuming the truth of an abstract idea,[194] and proceeds from this assumption to establish knowledge of particular things. For example, a rationalist might start with the idea that babies make dirty nappies; identify Ben as a

[186] See Hobbes comment in *Leviathan*, n 102, 658 that '… it was hard for men to conceive of those Images in the Fancy, and in the Sense, otherwise than of things really without us'. Hence his warning (136–7) that 'In Demonstration, in Councell, and all rigorous search of Truth, Judgement does all; except sometimes the understanding have need to be opened by some apt; and then there I so much use of Fancy. But for Metaphors, they are in this case utterly excluded. For seeing they openly professe deceipt; to admit them into Councell, or Reasoning, were manifest folly.'

[187] See *Westdeutsche Landesbank Girozonetale v Islington London Borough Council* [1996] AC 669.

[188] See *Re Charge Card Services Ltd* [1987] 1 Ch 50.

[189] See the judgments of Lord Hoffmann in *BCCI (No 8)* [1994] AC 214 and *Meridian Global Funds Management Asia Ltd v Securities Commission* [1995] 3 All ER 918.

[190] 'What should be absolutely clear, though, is that the process of deciding cases involves a much broader decision on the relative importance of theoretical and pragmatic considerations.' RWJ Hickey, 'Dazed and Confused: Accidental Mixtures of Goods and the Theory of Acquisition of Title', 368 *MLR* 2003 382.

[191] This is of course an oversimplification. In practice, elements of the two traditions are found throughout philosophy. See William James, 'Pragmatism' in *Pragmatism and Other Essays*, (Washington Square Press, 1968), 8: 'No one can live an hour without both facts and principles, so it is a difference …of emphasis.' In *Faith and Order; The Reconciliation of law and Religion* (Scholars Press, Atlanta, Georgia 1993), Harold Berman comments p 32 that 'a … characteristic of the Western legal tradition is the tension that has existed between its ideas and its realities, between its dynamic qualities and its stability, between its transcendence and its immanence.'

[192] Of course, a simple equation of rationalism with a priori thought is in some ways misleading; however, like many generalisations, it is a useful starting point.

[193] In 'Pragmatism', n 191, 9, William James goes on to caricature the 'tender-minded' rationalist as going by principles, intellectualistic, idealistic, optimistic, religious, free-willist, monistic and dogmatical, and the 'tough-minded' empiricist as going by facts, sensationalistic, materialistic, pessimistic, irreligious, fatalistic, pluralistic and sceptical.

[194] In classical thought, the basis of this assumed truth was divine; in medieval thought, such divine authority was mediated through the Catholic church. In his *Discourse upon Method* (Transl F E Sutcliffe, London, Penguin, 1968), Descartes offers a modern alternative based on reason.

baby; and thus deduce that Ben makes dirty nappies. In this way, rationalist thought operates 'top down'. In contrast, empirical thought rejects the primacy of abstract ideas.[195] It begins with concrete observations of particular things,[196] and aggregates these to form general ideas.[197] Thus the empiricist learns from the evidence of her senses that Ben makes dirty nappies; that Tilly, Naama and Jack do the same; and from this she induces the idea that babies make dirty nappies. Thus empirical thought operates *a posteriori*, or 'bottom up'.

A. Rationalism

The priority of ideas in rational thought is not merely methodological but principled. Abstract ideas are valued more highly than sensory data.[198] The physical and particular realm of sensory experience is mistrusted.[199]

It was seen in Section III that the dematerialisation of MMIs involves a movement from the physical and particular (in the form of the non-fungible

[195] Locke's equation of the new born human mind to a tabula rasa or blank slate is a rejection of innate ideas. See Locke, n 178, 84: '… there are no practical Principles wherein all Men agree; and therefore none innate.'

[196] See for example Locke, n 178, 680: 'Whereas, I truth, the matter rightly considered, the immediate Object of all our Reasoning and Knowledge is nothing but Particulars.'

[197] Locke, n 178, 373, 376, 504. See also 520: 'The whole extent of our Knowledge, or Imagination, reaches not beyond our own ideas, linked to our ways of Perception.' In *An Enquiry Concerning Human Understanding*, T Beauchamp (ed) (Oxford, Oxford University Press, 1999), David Hume argues that the senses are the foundation of knowledge. Sensation precedes thought (135), and impresses the mind more vividly (99). Knowledge is based on sensory experience (194, 196, 209), and extended by association (102, 115, 116) and analogy (165, 174, 175). Cartesian doubt concerning the evidence of the senses is unreal and unhelpful (199, 200). (On the last point, see also Locke, n 197, 46: 'If we will disbelieve everything, because we cannot with certainty know all things; we shall do much what as wisely as he, who would not use his legs, but sit still and perish, because he had no wings to fly.'

[198] '…the Theory of Forms … constitutes Plato's distinctive contribution to metaphysics.' Plato, *Phaedrus,* (Transl. Walter Hamilton, Penguin, London, 1973) Introduction, 17.

'Plato's Theory of Ideas or Forms, which reduced to its barest elements, is that the manifold and ever-changing phenomena of the world of sense are imitations or copies of eternal and absolute Forms, which alone have true reality, and to "participation" in which the sensible world owes such partial reality as it possesses.' Plato, *The Symposium* (Transl Walter Hamilton, Penguin, London, 1951), Introduction, 20. Plato contrasts the unchanging realm of the Forms (See *Phaedrus*, n 198, 52) which constitutes reality and may be the object of knowledge, available to the intelligence, with the changing world available to the senses, which may only be the subject of opinion (See *The Republic*, (Transl, Desmond Lee, 2nd edn, Penguin, London, 1975), 250, 251. The theory of the Forms rejects the particular in favour of the universal (*The Republic,* 224); this involves in turn a rejection of the objects of the senses (*The Republic*, 246), the only value of which is to recall the Forms. The senses are unreliable (*The Republic*, 370), and cannot be the basis of knowledge (*The Republic*, 278). Knowledge is properly acquired without reference to the senses, by deduction from principles (*The Republic*, 253, 254)

Like Plato, Kant distinguishes between phonomena (things which appear) and noumena (things which are thought). The latter are the foundation of ethics (See *Critique of Pure Reason* 2nd edn, 1787).

[199] In *Meditations,* n 87, Descartes mistrusts the world of the senses as the work of an evil spirit ('malin genie') (I 100): '… some deceiver … who constantly uses all his wiles to deceive

instrument) to the abstract idea of debt, recorded by (intangible) electronic record. This accords with the bias of rationalist thought.[200] Continental civil law was codified by rationalist thinkers in the early modern period, and this may in part explain the relative ease with which continental European markets have dematerialised their domestic securities.

B. Empiricism

However, common law rejects rationalist thought.[201] Common law commercial judges and writers regularly claim to know nothing about metaphysics, and to be the better for it.[202] There are strong conceptual links between English commercial law and empirical thought.[203] Hume's presentation of the empirical method in *An Enquiry Concerning Human Understanding* contains many of the themes subsequently developed in common law judgments and text books. These include in particular an emphasis on custom as the basis of knowledge;[204] a robust preference for practical experience[205] over abstract thought; distaste for metaphysics;[206]

me…' (II 103). 'I suppose therefore than all the things I see are false …' (II 102) and that the appearance from his window of men passing of the street may really comprise wax dummies moved by springs (II 109, 110). 'I shall now close my eyes, stop up my ears, turn away all my senses, even efface from my thoughts all images of corporeal things, or at least, because this can hardly be done, I shall consider them as being invalid and false …' (III 113). See also VI 154, 155.

[200] Walter Benjamin associates the movement from unique to fungible objects with abstraction and rationalist thought. 'To pry an object from its shell, to destroy its aura, is the mark of a perception whose 'sense of the universal equality of things' has increased to such a degree that it extracts it even from a unique object by means of reproduction.' 'The Work of Art in the Age of Mechanical Reproduction', in *Illuminations* (Fontana, London, 1992) p 217.

Dematerialisation involves the separation of the debt from the paper. It might be envisaged as a form of Cartesian dualism: see *Discourse upon Method*, n 194, 76: '… our soul is of a nature entirely independent of the body …'. See also *Meditations*, n 87, I 156.

[201] See for example, Hume, n 197. He argues that abstract reasoning is ill suited to the nature of human understanding (92). It is divorced from practice (88); powerless to discover cause and effect (110, 111); unfalsifiable (88, 205); and in the end, irrelevant (120, 206).

[202] For a famous example, see the many comments in Maitland's 'Trust and Corporation' in *Selected Essays* (Cambridge University Press, 1936) including the following (at 182): 'On juristic elegance we do not pride ourselves, but we know how to keep the roof weather-tight.' At 198 he refers to the 'petrifying action of juristic theory.'

[203] And there appears to be an inverse historic relationship between abstract philosophy and commercial and financial development. Compare the relative positions of England and Germany in the early modern period.

[204] 'All inferences from experience, therefore, are effects of custom, not of reasoning [122].

Custom, then, is the great guide of human life. It is that principle alone which renders our experience useful to us, and makes us expect, for the future, a similar train of events with those which have appeared in the past. Without the influence of custom, we should be entirely ignorant of every matter of fact, beyond what is immediately present to the memory and senses.' Hume, n 197, 121, 122. See also 129: 'This transition of thought from the cause to the effect proceeds not from reason. It derives its origin altogether from custom and experience. … This is the whole operation of the mind, in all our conclusions concerning matters of fact and existence'.

[205] 'The most lively thought is still inferior to the dullest sensation.' at 96.

[206] Hume, n 197, refers to '… the absurdity of all the scholastic notions with regard to abstractions and general ideas.' at 203.

and refusal to the exclude alternative principles which may inform particular instances. Hume's assertion that one effect may have different causes[207] accords with the common law practice of recording multiple judgments in a single case; it accords with the flexibility of the common law which Karl Llewellyn calls the Grand Style.[208] The famous jurisprudential debate between Hart and Kelsen may be understood as a debate between empiricism and rationalism.[209] When one considers the epistemology of common law, it may not be fanciful to compare pleadings to the bodily senses: the common law knows the commercial world through legal actions as a mind knows the world through the senses.

Anglo-Saxon philosophers and commercial practitioners have long attributed ethical value to empirical thought, which (like a democratically elected government) may be challenged when it is proved wrong by experience. Because a rationalist idea may not be dethroned or corrected by sensory falsification,[210] it is equated both with arbitrary power[211] and with folly.[212] There is a (dubious) tradition in empirical writing generally,

[207] 'In a word, I much doubt whether it be possible for a cause to be known only by its effect (as you have all along supposed) or to be of so singular and particular a nature as to have no parallel and no similarity with any other cause of object, that has ever fallen under our observation. It is only when two *species* of objects are found to be constantly conjoined, that we can infer the one from the other; and were an effect presented, which was entirely singular, and could not be comprehended under any known *species*, I do not see, that we could form any conjecture or inference at all concerning its cause. If experience and observation and analogy be, indeed, the only guides which we can reasonably follow in inferences of this nature; both the effect and cause must bear a similarity and resemblance to other effects and causes, which we know, and which we have found, in many instances, to be conjoined with each other. I leave it to your own reflection to pursue the consequences of this principle.' at 198.

Compare Sigmund Freud, 'Our imperative need for cause and effect is satisfied when each process has one demonstrable cause. In reality, outside us this is hardly so; each event seems to be over-determined and turns out to be the effect of several converging causes.' Sigmund Freud, *Moses and Monotheism* (transl Katherine Jones, Vintage, New York, 1939) at 137.

[208] Karl N Llewellyn, *The Common Law Tradition, Deciding Appeals* (Little, Brown and Company, Boston 1960). He attributes the success of the common law in effect to its empirical method: '… a court should seek to channel the impetus from the concrete.' (270).

[209] The author is grateful to Professor Lacey for this point.

[210] See Hume, n 197, 88: 'It is easy for a profound philosopher to commit a mistake in his subtile reasonings; and one mistake is the necessary parent of another, while he pushes on his consequences, and is not deterred from embracing any conclusion, by its unusual appearance, or its contradiction to popular opinion. But a philosopher, who purposes only to represent the common sense of mankind in more beautiful and more engaging colours, if by accident he falls into error, goes no farther; but renewing his appeal to common sense, and the natural sentiments of the mind, returns into the right path, and secures himself from any dangerous illusions.'

This anticipates Popper's argument that truth and therefore open society depends on empirical enquiry, which predicates the possibility of falsification.

[211] In contrast, see Blackstone's equation of common law with civil liberty and of civil law with political oppression: Blackstone, n 14, I 66, 67, 122, 123.

[212] Hobbes rejects scholastic writing as '… nothing else for the most part, but insignificant Traines of strange and barbarous words …', Hobbes, n 102, 701. Hume, n 197, mockingly equates Schoolmen to robbers in the forest (91).

and in commercial common law writing in particular, of deriding the unreal[213] civilians as both dangerous[214] and unsuccessful. Thus (it is suggested in such writings), unassisted by the light of common sense and the bracing friction of commercial practice, rationalists stumble around in the darkness,[215] remote from reality,[216] asserting arbitrary power and lapsing into extravagant errors.[217]

There may be an unconscious equation between the removal of paper upon dematerialisation, and the removal of the empirical base in the conceptual structure of financial law and practice. To common lawyers, this may gives rise to a nagging sense that the step is morally, cognitively and even politically unsafe.[218] Hence perhaps the stubborn sense, in spite of argument and evidence to the contrary, that dematerialisation raises intractable legal and operational difficulties.

IX. PSYCHOANALYTIC ASSOCIATIONS

One of the haunting curiosities of financial law is that it speaks in the language of love. It is full of puns, in which the first meaning is a term of financial law, and the second meaning concerns interpersonal attachment. For example: bond, share, participation, commitment, trust. For the author at least, a realm of emotional experience is obscurely evoked but not clearly presented. It is like something almost forgotten.

This may be unsurprising in view of the central technique of financial law. This is to treat an obligation as if it were a thing, which can be subject to property rights and readily transferred from person to person. Of course, the reification of debt is the basis of market capitalism. However, it has

[213] See Hume, n 197, 88: 'On the contrary, the abstruse philosophy, being founded on a turn of mind, which cannot enter into business and action, vanishes when the philosopher leaves the shade, and comes into open day; nor can its principles easily retain any influence over our conduct and behaviour.' See also James, n 191, 13: 'I will dwell a little longer on that unreality in all rationalistic systems by which your serious believer in facts is so apt to be repelled.' (214).

[214] See Hobbes' portrayal, n 102, of those affected by classical thought as regicidal (369) and hydrophobic (370).

[215] See Hume, n 197, 206: 'These principles may flourish and triumph in the school; where it is, indeed, difficult, if not impossible to refute them. But as soon as they leave the shade, and by the presence of the real objects, which actuate our passions and sentiments, are put in opposition to the more powerful principles of our nature, they vanish like smoke…'.

[216] Hume, n 197, 203, discusses '…the absurdity of all the scholastic notions with regard to abstraction and general ideas.' See also William James, n 191, 14: 'Truly there is something a little ghastly in the satisfaction with which a pure but unreal system will fill a rationalist mind.'

[217] See Hume's mockery of the doctrine of infinite divisibility, Hume, n 197, 204. Such abstract thought addles the mind: 'Reason here seems to be thrown into a kind of amazement and suspense, which, without the suggestions of any sceptic, gives her a diffidence of herself, and of the ground on which she treads' (205).

[218] With dematerialisation, it is perhaps unconsciously felt that we are somehow not safe in our beds, and the continentals are behind it.

encountered both conceptual resistance from lawyers[219] and ethical resistance from sociologists and others.[220] The essence of the objection is always the same: an interpersonal relation (of obligation) is wrongly characterised as a relation between a person and a thing. The inter-subjective element is lost.

Jacques Lacan wrote about the costly tendency, revealed in psychoanalysis, to deny inter-subjectivity. At the mirror stage, when the infant differentiates herself from the mother, she is distressed to learn that she does not control her mother. In her first potentially inter-subjective relationship the infant experiences powerlessness and therefore pain. In order to avoid this pain she turns to objects, which she may control. As an adult, she continues to reach for objects, with a desire that is fierce because it is always disappointed. It is disappointed because it is not for objects, but for inter-subjectivity. Thus one aspect of materialism is the displaced desire for human relationship.

Professor Jeanne Schroeder has applied Lacanian theory to the financial markets.[221] She argues persuasively that the stubborn and troubled attachment of market participants to paper instruments is a form of Lacanian desire.

X. RELIGIOUS ASSOCIATIONS[222]

In its historical development, financial law drew heavily on Canon law,[223] and also on the practices of international religious communities.[224] Many great judges were religious men, and many religious leaders were lawyers.[225] It has been argued that the historic development of the common law is due in part to the rise of religious movements in England.[226] Strong cultural links

[219] An example of this is the protracted debate surrounding the concept of PRIMA in the conflict of laws: see C Bernasconi, *The Law Applicable to Dispositions of Securities Held through Indirect Holding Systems* (Hague Conference on Private International Law, 2001).

[220] See for example, Karl Marx, *Capital*, v 1, ch 1.

[221] See for example Schroeder, n 2.

[222] 'Law's space has not entirely shed its sacred origins: Kahn, n 14, 61. Also see Bernan, n 190, who seeks 'to show that despite … tensions the two [law and religion] stand in a dialectical relationship with each other and cannot maintain their vitality independently of each other.' (x).

[223] Berman, n 190, shows how Roman law entered the modern legal world via cannon law (27, 32). See also *Blackstone*, n 14, II 441: 'Equity … was laid down by the clerical chancellors.'

[224] See eg *Blackstone*, n 14, IV 412: 'Trade, or foreign merchandise, such as it then was, was carried on by the Jews …' This theme is developed (controversially) by Werner Sombart, *Jews and Modern Capitalism,* (Transl M Epstein, Fisher Unwin, London 1913). Weber, Tawney and Berman, n 190, alike explore the involvement of Puritans in commerce.

[225] Berman, n 190, comments that Calvin was a lawyer (105) and Hale a Puritan (121, 122).

[226] See Berman, n 190, 116: 'Some external connections are obvious. Puritanism was undoubtedly the spark which ignited the Civil War, which in turn led eventually, after many upheavals, to the victory of Parliament over the Crown and of the common law courts over its rivals.'

have been shown between commercial practice and Protestantism[227] (but not Catholicism[228]). Empirical philosophers[229] and common lawyers[230] have developed their arguments in overtly anti-Catholic terms.

Conceptual links are identified between the commercial bargain and the religious covenant.[231] The rise in the eighteenth century of the highly developed conceptual structures of modern financial law coincided with the decline of theology in the prevailing discourse. Debate about money replaced debate about God, largely in the same terms.

A poetic example of such religious, and indeed mythic, associations is provided by an ancient text which in many ways in foundational in Western culture. [232] This is the book of Genesis, and in particular two linked passages, one drawn from the story of creation, and one relating to the covenant made by God following the flood.[233]

The first passage is Genesis 1.14. On the fourth day, God created the sun, moon and stars.

> And God said, Let there be lights in the firmament of the heavens to divide the day from the night; *and let them be for signs*,[234] and for seasons, and for days, and years.[235]

[227] In *The Protestant Ethic and the Spirit of Capitalism*, (Routledge, London, 1992), 172, Max Weber discusses '… the religious valuation of restless, continuous, systematic work in a worldly calling, as the highest means to asceticism, and at the same time the surest and most evident proof of rebirth and genuine faith, must have been the most powerful conceivable lever for the expansion of that attitude toward life which we have here called the spirit of capitalism.' See also 175.

Tawney, 204, discusses '[t]he identification of the industrial and commercial classes with religious radicalism …', and the development of this identification into '… the theory, later epitomized by Adam Smith in his famous reference to the invisible hand, which saw in economic self-interest the operation of a providential plan.' (195).

Blackstone's rejection of 'Roman' civil law is overtly anti-Catholic, and his celebration of common law as an independent spirited rejection of the 'Roman' system, attracting the accusation of heresy, is associated with a celebration of Protestantism in England. See for example I 20, 21.

[228] R H Tawney, *Religion and the Rise of Capitalism* (first published 1926) (Penguin, London, 1990), 206.

[229] Hume, n 197, famously ends his *Enquiry* by consigning the works of the Schoolmen to the flames (211). See also Hobbes' argument attributing the rejection by the Protestant States of Catholicism to the effect of '… the canting of Schoolemen' (Hobbes, n 102, 115) in … bringing of the Philosophy, and doctrine of Aristotle into Religion … from whence there arose so many contradictions, and absurdities, as brought the Clergy into a reputation both of Ignorance, and of Fraudulent intention; and enclined people to revolt from them, either against the will of their own princes, as in *France* and *Holland*, or with their will, as in *England*.' (182). Hobbes, n 102, compares the Papacy with the kingdom of faries (712, 3) and with the Greeks with demons (639).

[230] See Blackstone, n 14, I 17–18 and III 99, 10.

[231] See Berman, n 190, pp 202–206.

[232] The author is grateful to Professor Robert Reiner, inter alia for his help in relation to this section of this paper.

[233] In Jewish tradition, there are strong links between the two passages, as Noah is equated to Adam, being offered a new beginning.

[234] Author's italics.

[235] Genesis 1.14.

The word for 'sign' in Hebrew is 'ot'. These lights *represent* day and night and the other periods; it follows that they do not *constitute* them. The day and night for which the lights serve as signs remain outside creation. They exist only as the absent signified, outside the sensory world, in the mind of God. This accords with the status of the debt instrument before the legal fiction took effect. These represented the debt, but did not constitute it; the debt existed only as a chose in action, outside the sensory world, at law.

Thus in Genesis 1.14 the lights in the sky serve as the visible signs of the intention of God. After the flood, the rainbow is added to these as the visible sign of God's covenant.

> And God said, This is the token of the covenant which I make between me and you and every living creature that is with you, for perpetual generations: I do set my bow in the cloud, and *it shall be for a token*[236] of a covenant between me and the earth.[237]

The same word, 'ot', again describes the signifying function.[238]

God assures us that we may rely on God's covenant because of its visible sign, the rainbow. The financial markets are asked to rely on issuers' covenants, and are uneasy at the withdrawal of its visible sign, the paper.

Part of the covenant set out in the second of the linked passages (in Genesis 8.22) is that:

> While the earth remaineth, seedtime and harvest, and cold and heat, and summer and winter, and day and night shall not cease.

There are strong poetic associations between 8.22 and 1.14, in both the Hebrew and Authorised English versions. The imagery and rhythm of the two passages are so alike that when one reads either, one thinks of the other. These textual links serve to associate the ideas in the two passages. Thus the signifying function of visible creation (1.14) is linked with God's covenant (8.22). Genesis 8.22 says that these two will continue together.

It does not exclude the possibility that they may also cease together. Although traditionally the covenant is taken to mean that the flood will never come again, 8.22 does not say so. Its promise is qualified as to time. '*While the earth remaineth*...'[239] the distinctions of seedtime and harvest, and cold and heat etc shall not cease. However, the covenant would permit the cessation of these distinctions with the cessation of the earth.

[236] Author's italics.
[237] Genesis 9.12 and 13.
[238] 'And the Jews had that quality which St. Paul noteth, *to look for a sign*.' Hobbes, n 102, 507.
[239] Author's italics.

We are taken back to the opening passages of Genesis, where a chaotic earth 'without form and void', predates the process of creation.[240] Creation proceeds by the imposition of order through a process of distinction.[241] The separate parts of visible creation are then able to bear meaning, and to refer to intention of God beyond themselves.[242] Distinct, visible things have meaning, and (according to the covenant) will continue to do so as long as they continue. However, if distinct visible things disappear, distinction and meaning themselves may collapse. We may be left without God, flooded out.[243]

The book of Genesis is our foundational myth, and its influence on our thoughts is not reasoned nor necessarily conscious, but nonetheless deep. Hence perhaps the anxiety surrounding the loss of the physical instrument. The anxiety may give way to dread when it is learned that dematerialised MMIs are not only intangible, but also fungible (or, in the language of the St James' bible, without form and void).

[240] Genesis 1.2.
[241] Genesis 1.4; 1.6; 1.7; 1.9; 1.14; 1.18.
[242] The Hebrew original for the phrase 'without form and void' is tohu v'mohu. I believe this might also be translated as 'meaningless'.
[243] It is understood that Clearstream operates from a vault which is tanked, even though it is some 200 miles from the sea.

Commentary on 'The Dematerialisation of Money Market Instruments'

GUY MORTON

I. INTRODUCTION

DR BENJAMIN'S STIMULATING chapter covers a wide range of issues. This note does not attempt to cover all such points, partly through constraints of space and partly (especially in relation to Part IV of Dr Benjamin's chapter) because of the limitations of the writer.

This note inevitably tends to focus on points on which the view, or the emphasis, of the writer differs from that of Dr Benjamin. This should not be taken to imply disagreement with Dr Benjamin's assessment of the decision not to attempt to replicate the traditional concept of negotiability in the context of dematerialised money market instruments. Even if this was possible, the writer agrees that it would not have been desirable. He also agrees with Dr Benjamin's broader conclusion that English law relating to the transfer of securities does not deal adequately with issues of third party claims and priorities and needs unified new rules in this area.

II. CAN INTANGIBLES BE NEGOTIABLE?

Dr Benjamin argues that intangibles cannot be negotiable, because they cannot be subject to possession. As a proposition based on the traditional conceptual framework this is clearly correct; but is it merely an accident of history (and therefore reversible given sufficient authority to change the law, eg by statute) or the product of some more essential element of negotiability? Dr Benjamin appears in places to be somewhat ambivalent about this; for example she states[1] that 'it is undesirable [but not, therefore, impossible?]

[1] Section IVCb.

that intangibles should be categorised as negotiable instruments, because the result of applying the many rules predicating paper to an intangible would be highly uncertain'.

Footnote 16 of Dr Benjamin's chapter, quoting from the 1999 Bank of England paper, identifies four main legal features of negotiable instruments, namely that—

(a) they can be transferred solely by physical delivery;
(b) they provide the holder with unchallengeable title;
(c) they provide certain benefits in the event of a court action; and
(d) they may include rights against two or more parties.

The question is whether these are all *essential* features, in the sense that something that does not possess all four cannot accurately be described as negotiable. Dr Benjamin notes that security of transfer (freedom from adverse claims) is the chief advantage of negotiability; if a dematerialised instrument can be given this (eg by statute, for example under section 207 of the Companies Act 1989), why cannot it be described as negotiable?

Some support for the view that it is in principle possible for a dematerialised instrument to be negotiable is given by the draft Negotiable Instruments Act in Appendix A to the Jack Committee Report,[2] section 84 of which expressly provides for dematerialised negotiable instruments.

On reflection, however, it is less clear whether this draft section conclusively supports the application of the concept of negotiability to dematerialised instruments. Indeed it can be argued that it actually supports the contrary view, that the true category of negotiability is indeed limited to instruments capable of actual delivery. This is because the draftsman feels it necessary to apply the techniques of cross-reference and 'deeming', implying that dematerialised instrument will have, at most, only 'quasi-negotiability'.

It is, perhaps, a question of semantics whether, if a statutory framework for perpetuating features of traditional paper negotiable instruments were to use the terminology of negotiability, the result would truly be a new category of negotiable instrument. The more important question is whether the use of the old terminology would confer any benefit. The present writer agrees with Dr Benjamin that it would not. As Dr Benjamin states in the section of her chapter dealing with security of transfer:[3]

> the general principles of English law do not deal adequately with the question of third party claims. Security of transfer is reduced, and legal risk is increased, as the outcome of any dispute may depend arbitrarily on the operational detail of the parties' custody arrangements. Generally, this reflects the

[2] *Banking Services: Law and Practice—Report by the Review Committee* (February 1989, Cm 622), pp 301, 302.
[3] Section VID.

need for a unified law of priorities, which does not differentiate between legal and equitable interests.

Measured against the objective of a unified law of priorities, the perpetuation of the old conceptual framework—itself largely the product of the historical forces described by Dr Benjamin—would tend to increase complexity without any obvious compensating benefit. It would therefore be defensible only if the class of instruments to which it relates are in practice treated in a special way which requires or justifies distinguishing them from other securities. There appears to be no evidence to suggest that this is the case—indeed the general acceptance by market participants, in the responses to consultation, that the normal framework of the Uncertificated Securities Regulations provided an acceptable basis for dematerialised money market instruments suggests the contrary.

III. FUNGIBILITY

As an additional ground for her conclusion that dematerialised money market instruments cannot be negotiable, Dr Benjamin argues that they are fungibles and that fungibles cannot be negotiable. This argument is expounded in an interesting and wide-ranging section, which raises points that go beyond the specific question of negotiability.

Dr Benjamin's argument that fungibility is incompatible with negotiability is initially expressed in the proposition that 'fungibles cannot ... be individually possessed'.[4] It was pointed out in the discussion at the seminar that this seems too wide, since some fungibles—banknotes, for example—clearly can be individually possessed. In acknowledgment of this point, Dr Benjamin modifies the proposition to state[5] that 'undivided fungibles cannot be individually possessed'.

In fact much of the discussion in this section of Dr Benjamin's chapter relates to issues of ownership rather than of possession. This appears to be because, as Dr Benjamin puts it,[6] 'the law of possession is less developed than the law of property'. The implication seems to be that the degree of identification or segregation required for possession must at least correspond with that required for property rights, so that the law on property rights is indirectly relevant.

[4] Section V introduction. The word 'individually' is clearly important, since Dr Benjamin has already concluded that dematerialization destroys negotiability by removing the possibility of possession and delivery. In other words the sequence is 'dematerialized MMIs cannot be possessed; and even if they could be possessed, they could not be *individually* possessed.'
[5] Fn 72.
[6] Section VD.

Dr Benjamin accordingly goes on to consider the status of securities as undivided assets. She appears to agree with Professor Goode[7] that registered securities are coproprietary interests in an undivided asset,[8] but to differ from him, at least in terminology,[9] in referring to such interests as fungibles. The difference centres on the argument that there can be differentiation between parts even of an undivided asset,[10] and that it therefore remains possible to describe differentiated parts of such as asset as fungible with each other.

In what follows, the present author comments on two issues raised by these views: first, whether there can be individual property rights in parts of an undivided asset; and secondly, whether Professor Goode and Dr Benjamin are correct in regarding the share capital of a company as an undivided asset.

A. Property Rights in Parts of an Undivided Asset

Discussion of the first question starts from the familiar rule that for a property right to exist, its subject matter must be identified. In the case of what Dr Benjamin calls 'ex-bulk fungibles', the subject matter of the possible property right is partially identified, whereas with 'generic fungibles' it is not identified at all; but in either case, the classical analysis would be that a property right cannot exist unless its subject-matter is *fully* identified.[11] In other words, one would expect the answer to Dr Benjamin's question:[12]

> whether [property rights] ... require their subject assets to be allocated, so that a particular property right attaches to a particular asset

to be 'yes'.

Dr Benjamin argues that the fact that a floating charge confers property rights 'supports the proposition that property rights need not attach to particular assets, and therefore that ex bulk fungibles may be subject to property rights'. This argument seems to the present writer to be misguided. While the peculiar nature of the rights conferred by a floating charge will doubtless continue to be a source of debate, there is at least no difficulty about identifying the property in which the rights conferred by a floating charge subsist

[7] Roy Goode, 'Are Intangible Assets Fungible?', ch 7 in Birks and Pretto (eds) *Themes in Comparative Law in Honour of Bernard Rudden*, (Oxford University Press 2002), 101, cited by Dr Benjamin in n 36 and subsequently.
[8] My comments below in general follow Professor Goode in concentrating on shares. See n 21 below for a brief consideration of debenture stock.
[9] N 87: 'The author hopes that [Professor Goode] will forgive her for persisting in using the term 'fungible' in a wider sense, to include parts of undivided assets'.
[10] N 87.
[11] As Professor Goode points out, in the chapter cited in n 6 above, this does not prevent property rights, including coownership rights, existing in a bulk—the bulk as a whole being fully identified.
[12] Section VC.

(namely the whole of the property, or of the relevant class of property, of the chargor). Whether property rights can subsist in only part of an unallocated pool is a different question. If therefore the implication of Dr Benjamin's observation that 'the want of attachment for ex bulk fungibles is internal to the pool, where the want of attachment for floating charges is coterminous with the pool' is that the nature of the conceptual difficulty is the same in both cases, it is surely incorrect. Both the notion of an attached property interest in unidentified property and that of an unattached property interest in identified property raise conceptual difficulties. In the first case, most lawyers have traditionally found the difficulties insuperable; in the second, the common law at least has succeeding in developing a floating charge concept which accommodates the difficulty to its own satisfaction (though it continues to baffle lawyers from other traditions—and perhaps some common lawyers too!).

Dr Benjamin states that 'it is clear that where the fungibles are ex bulk, the subject matter of the trust is sufficiently certain' and that, accordingly, 'it is clear that property rights can exist in relation to ex bulk fungibles'. By this she appears to mean that property rights can exist in assets forming part of a fungible bulk notwithstanding absence of severance or allocation. However, in the same passage she observes that 'the precise structure of the property rights of beneficiaries is still open to debate' and that 'the options [for the possible form of such rights] available include (and may be limited to) co-ownership and charge'. Both co-ownership and charge, in their application to a fungible bulk of assets, are concepts which avoid any problem of identification (in the trust context, that of certainty of subject matter) by creating proprietary interests over the whole bulk. If therefore (as Dr Benjamin acknowledges may be the case—presumably because of the difficulty of identification?) these are the only options, it appears to follow that it is *not* clear, after all, that there can be property rights in ex bulk fungibles.

The question of whether full identification is a pre-requisite for the existence of individual property rights in securities therefore comes back to the decision in *Hunter v Moss*. A full discussion of that case is beyond the scope of this note. The criticisms of Professor Hayton and others appear to the present writer to have considerable force; and while there are undoubtedly counter-arguments,[13] they do not seem to provide a clear and coherent alternative to the structure of co-ownership of the entire pool propounded by Dr Benjamin in her footnote 130, which, therefore, the present author continues to recommend as the preferable basis of documentation in practice.[14]

[13] See in particular Dr Worthington's interesting article at [1999] *JBL* 1, cited by Professor Goode.

[14] Under the current law. The fact that it is possible to achieve a workable model using existing legal concepts does not mean that it would not be preferable to create a unified legal framework for holding and dealing in securities under modern market arrangements.

B. Shares as Units of a Single Undivided Asset

This note would however be incomplete without some discussion of fungibility in relation to shares. This is an area where the writer—with some trepidation—finds himself differing from the views both of Dr Benjamin and of Professor Goode.

Professor Goode propounds the view that shares are merely coownership interests in a single asset—namely the share capital of the relevant company—and are therefore not fungible at all.

Dr Benjamin[15] does not follow Professor Goode completely, on the ground that differentiation between shares does occur in the context of equitable tracing claims. On the other hand, she seems to concur with Professor Goode's basic thesis by supporting the view that shares are 'fractional interests in the undivided capital of the issuer'.[16]

This note will not pursue the issue of the relevance of equitable tracing claims, but will concentrate on the more basic question raised by Professor Goode.

The starting point of Professor Goode's analysis is that:

> fungibility involves a choice between legally interchangeable units, whether from a wholly unidentified source or from a larger identified **collection or bulk**. It therefore presupposes that **the asset said to comprise the units or bulk** is of a kind that is divisible into separate units capable of being separately owned. In other words, there must be a **collection or mass** from which individual units can be **detached and** disposed of separately, so that the purchaser of the segregated units **has no interest, or ceases to have an interest**, in the remainder of the **collection or bulk**. Coownership of a **single, indivisible asset** is therefore incompatible with the concept of fungible units …

Later he restates the principle as follows:

> In **every** case the principle is the same: *the question of identification through segregation does not arise unless the property the subject of the enquiry is divisible into units capable in law of being owned separately from* **the property of which they previously formed part.**

and later still, as follows:

> The fundamental requirement of interchangeability means that *the property the subject of the enquiry must be susceptible of being divisible into separate units capable in law of being separately owned.*

[15] N 87.
[16] Section VAb.

The italics are Professor Goode's; the emphasis by bold type is added. The purpose of the addition is to suggest that Professor Goode describes two distinct models or sets of concepts, but does not fully articulate one of them, and formulates his conclusion on the basis of the other. In the view of the current writer, this leads to a conclusion which is, at least, too categorical.

The two 'models' referred to above can be called the 'bulk' model and the 'collection' model. The key relevant features of the bulk model are the following:

(a) the bulk is, or can sensibly be regarded as, a single asset;
(b) the units cannot sensibly be regarded as independent assets while they remain part of the bulk (and accordingly must be divided from the bulk before they can be so regarded);
(c) while a unit remains part of the bulk, the owner of a unit has an interest, not only in 'his' unit, but in all the other units which form part of the bulk.

The key features of the 'collection' model are less clearly articulated. However, the writer suggests the following features, some (but perhaps not all) of which appear to be supported by Professor Goode's terminology:

(a) the units can sensibly be regarded as independent assets while they form part of the collection;[17]
(b) the collection itself may be merely a descriptive term referring to the class of units taken together, or (possibly) may itself also be an asset;[18]
(c) the holder of a unit does not have an interest in other units.[19]

Professor Goode's conclusion is, however, drawn exclusively from the 'bulk' model. Having shown convincingly that fungibility under the 'bulk' model depends on the possibility of units being detached from the bulk, he is of course able to reach the further conclusion that units which are incapable of being detached from the bulk cannot be fungible. If therefore the 'bulk' model is the only, or the correct, model for share capital, it indeed follows that shares cannot be regarded as fungible. But is this the case? It is far from clear to the present writer that it is.

It is worth observing, in the first place, that if the 'single asset' analysis of share capital is correct, then the terminology commonly used by judges and

[17] This is not stated in terms but would seem to be implicit in the use of the term 'collection' as an alternative to 'bulk', and perhaps also in the phrase 'the asset said to comprise the units or bulk'.
[18] 'The asset said to comprise the units'?
[19] The phrase 'has no interest, or ceases to have an interest' appears to imply that in some cases the holder of one unit never has any interest in the others.

practising lawyers can at best be described as imprecise. In many cases the discrepancy between the language and [what on this view is] the reality is dramatic. Consider, for example, the following passage from *Crerar v Bank of Scotland* (1921) SC 736 (Court of Session; approved HL (1922) SC 137):

> When the shares of a joint stock company have been fully paid-up, a share of any particular denomination is, in the ordinary case, of the same pecuniary value as any other share of that denomination in the same company. It does not, however, follow that a borrower who transfers such shares to a lender in security of cash advanced has no interest to insist, and no right to expect, that the identical shares so transferred shall be retained and held for his behoof, and shall be restored to him upon his repaying all that he owes to his creditor. In the absence of some express or implied agreement to the contrary … the lender's duty … is to retain and hold on account of the borrower the identical shares which he received … . Moreover, although the point is hardly a practical one, a borrower would be entitled to say that he preferred to receive back his own shares, the title to which he knew to be good, rather than other shares, the title to which might turn out to depend on a forged transfer.

And, in England, *Ellis & Company's Trustee v Dixon-Johnson* [1924] 2 Ch 451:

> It would be absurd to insist on a retransfer of the identical shares mortgaged when other shares of the same nature are available.

On Professor Goode's analysis, such an insistence would seem to be not merely absurd as a practical matter (which was clearly the judge's meaning in the context), but absurd as a matter of logic—impossible and meaningless.

Other implied results of Professor Goode's analysis also have a surprising appearance. Take the question of voting rights. According to the analysis, the voting rights of shareholder A and shareholder B form part of a single asset which they hold in common with each other and the other shareholders. Is it not then rather odd that they can exercise these common rights in opposite ways?

The share capital which, on Professor Goode's analysis, is the only asset in which property rights subsist also has some unusual features. For most of its history it could not be owned (or at least lawfully owned) except through coownership. Moreover, at least on one view of the applicable rules of private international law, proprietary issues in respect of the coownership interests comprised in it can be governed by different systems of law.[20]

[20] This will be the case for a company which has a branch register, if it is correct to regard the location of the branch as fixing the *situs* of the shares recorded on it. See the cases on the *situs* of shares cited in Dicey & Morris, *The Conflict of Laws,* (13th edn) para 22–044 and the judgments of the Court of Appeal (in particular Auld LJ) in *Macmillan Inc v Bishopsgate Investment Trust plc (No 3)* [1996] 1 WLR 387.

These points do not, of course, refute Professor Goode's analysis—it would not be the first time that judges and practitioners, and even the legislature, have used 'common sense' language in a manner at variance with the legal reality. But they do give one pause.

Moreover, the co-ownership for which Professor Goode argues in this context seems to differ in an important respect from co-ownership as encountered elsewhere in relation to intangibles. The difference relates to the third of the elements set out in the text associated with footnote 19 above (that is, the interest of one holder of 'units' in the 'units' of another) and consists in the fact that the holder of a co-ownership interest in an intangible generally cannot enforce his rights without the cooperation of another person (either the other co-owners or a person holding the relevant property on behalf of the co-owners collectively). This becomes clearer when the analysis is extended to debts. Professor Goode observes that 'there can be no objection, on the ground of lack of identification, to the assignment of £100 forming part of a bank deposit of £500. There is only one asset, which following the assignment is in the co-ownership of the assignor and the assignee. It is not possible for an assignee to acquire part of the deposit without becoming a coowner of the entire deposit'. This is true; but there is an important distinction between a debt such as a bank deposit, the terms of which make no provision for partial assignment, and one such as a holding of debenture stock, the terms of which confer a right of subdivision. The partial assignee of a £500 deposit needs the concurrence of the assignor in order to enforce the debt against the debtor. With a transfer of debenture stock, the case is different. The transferor and the transferee are recognized by the debtor (the issuer) as independent holders[21] of their respective amounts of stock. Similarly a shareholder does not need the concurrence of other shareholders to enforce rights conferred by his shares.

This suggests that, at least in the case of intangibles which are things in action, a partial assignment which is made under a right of subdivision conferred by the terms of the thing in action itself may be regarded as creating a separate thing in action, in a way which is not true of a partial assignment which is only a matter between the parties to the assignment; and that, for so long as this separate thing in action either continues to be held by the assignee or is further assigned in a way which does not involve either further subdivision or consolidation with another like thing, it may be not only natural, but accurate, to speak of it as the 'same' asset and as an asset separate from the assets constituted by other holdings.

[21] Though it is true that the debenture stock trust deed may include restrictions on the ability of an individual stockholder to sue for the debt or enforce his rights in other ways such as presenting a petition for the winding up of the issuer. In that respect, debenture stock occupies something of an intermediate position between shares and debts such as deposits, the terms of which do not provide for subdivision. Nevertheless, the writer's view is that the right of subdivision separates both shares and debenture stock from simple debts for the purposes of the issues here under consideration.

Accordingly the writer[22] is not convinced by Professor Goode's contention that the 'repurchase' leg of a sale and repurchase agreement must of necessity relate to 'the very same' asset as the purchase leg. He continues to believe that judges may still look to a distinction between the identical securities transferred and equivalent securities as a relevant factor in characterising the transaction,[23] and that it is therefore prudent to draft against that possibility.

In the view of the writer it is strongly arguable that, where the terms of issue of a security confer a right of subdivision, subdivision of a holding is equivalent to severance in the case of a bulk of tangible assets (the bulk here being the holding immediately before subdivision, rather than the entire issue). This explains the language used by courts in cases such as those quoted above (and also by critics of the decision in *Hunter v Moss*). It also implies that Dr Benjamin is mistaken in regarding fungibility as an additional ground for the conclusion that dematerialised money market instruments cannot be negotiable. This is because that conclusion depends on the proposition that securities are undivided fungibles.[24]

Dr Benjamin also cites balance sheet treatment, legal analysis, historic development and the term 'share' as indicators of the nature of a share as 'fractional interests in the undivided capital of the issuer'. To comment briefly on these in turn—

(a) *Balance sheet treatment*: If this refers to the fact that the interests or claims of shareholders are entered on the balance sheet as a single aggregate item of 'share capital', it seems difficult to draw any reliable inference from this, when one recollects that other items which are clearly collections of separate liabilities (for example trade creditors) are similarly aggregated.

(b) *Legal analysis*: Dr Benjamin refers to *Ind's* case.[25] That case involved an employee of a company (Mr Ind) who had agreed to take a transfer of nil-paid shares as a nominee of the company. Through a mistake, the transfer to Mr Ind identified (by their numbers) shares which were not registered in the name of the transferor, but the transferor did hold the appropriate number of shares (with other numbers—one digit being different). The company

[22] Who should perhaps declare an interest: as one of those 'repo specialists' who participated in developing the detailed equivalent securities terminology now used in repo and other title transfer financing documentation, he is naturally reluctant to conclude that it is misconceived. Readers will doubtless take this into account in assessing the arguments in this section.

[23] So that, in particular, an undertaking to retain the particular securities transferred in a separate designated account and not to deal with them save on the occurrence of a default would be regarded as an indication that the true character of the transaction was that of a charge rather than an outright transfer.

[24] See Section IIIB above.

[25] *In re International Contract Company, Ind's case* (1872) 7 Ch App 485.

having gone into liquidation, Mr Ind sought to argue that he was not properly included on the list of contributories. Not surprisingly, the court had no sympathy for this argument. The two judges in the Court of Appeal in Chancery disposed of it in virtually identical terms:

> The substance of the transaction was that [Mr Ind] meant to be on the list for fifty shares; he was on the list for fifty shares, and the creditors and other persons interested have a right to hold him to that. (James LJ)
>
> I think the substance of the matter is that Mr Ind has agreed to take fifty shares as between him and the creditors of the company; he has been registered with his own consent as to these shares, and therefore he is properly on the list. (Mellish LJ).

Both judgments refer to the purpose of numbering shares being 'the tracing of title to the shares', and Mellish LJ states that 'one share, an incorporeal right to a certain portion of the profits of the company, is the same as another and … Share No 1 is not distinguishable from share No 2, in the same way as a grey horse is distinguishable from a black horse'. To the present writer, these comments are at least as indicative of a view that shares are separate, but interchangeable (and therefore fungible) assets as of the analysis suggested by Dr Benjamin and Professor Goode.

(c) *Historic development; name of shares*: These factors are double-edged, when one reflects on the case law[26] which establishes that statutory incorporation brought about a decisive shift in the judicial analysis of the relation between a company and its shareholders. The present writer is more swayed by the comment of Professor Gower that 'the word 'share' has become something of a misnomer, for shareholders no longer share any property in common; at the most they share certain rights in respect of dividends, return of capital on a winding up, voting, and the like.'[27]

Finally it is worth alluding to the observation made by Dr Benjamin in the introductory part of her chapter,[28] that 'it is at least arguable that the whole of the subject-matter of financial law is internal to financial law … See *Meridian Global Funds Management Asia Ltd v Securities Commission* [1995] 3 All ER 918, per Lord Hoffmann 922: 'There is in fact no such thing as the company as such, no 'ding an sich', only the applicable rules'.' This suggests that Professor Goode's analogy of shares in a horse should be

[26] In particular the *Macaura* and *Salomon* cases.
[27] Gower, *Principles of Modern Company Law,* 6th edn, p 300.
[28] N 14, 3rd para.

viewed with caution. A horse is a 'ding an sich' and it is obvious that it must be treated as an asset of a different order from any legal interest in it; but shares and share capital both owe their existence to what Lord Hoffmann calls 'the applicable rules'. Whether shares are to be viewed as interests in share capital or share capital as an aggregation of shares arguably depends on which end of the telescope the observer chooses to view the matter through. Most judicial observers appear to have viewed it through the 'shares as assets' end.

This discussion is very far from covering all the issues arising from Dr Benjamin's discussion of fungibility; but it is already much longer than was necessary merely to express scepticism about Dr Benjamin's use of fungibility as an additional reason for denying the possibility of dematerialised negotiable instruments. By way of apology or mitigation, the subject is clearly important, and the discussion suggests that it remains in need of clarification.

IV. NATURE OF TRANSFER OF REGISTERED SECURITIES

In Part III of her chapter Dr Benjamin argues that the legal basis for the transfer of securities in CREST is novation, and that the reason why a CREST transfer results in the transferee acquiring the security transferred free from equities is that 'the asset of the transferor differs from the asset of the transferor or other predecessor in title.'[29] The present writer finds this analysis difficult to accept.

The starting point in Dr Benjamin's reasoning on this point appears to be the observation that:

> The term 'transfer' is not a legal term. It refers to a transaction, the economic result which is that one person ('the transferee') holds the same asset, or an asset equivalent to the asset, previously held by another person ('the transferor'). There are a number of different legal techniques of transfer. Assignment, negotiation, and novation are relevant to this discussion.

The premiss in the first sentence is debatable. No doubt it is true that 'transfer' is not a legal term in the sense of a word which is always, or nearly always, used in a technical legal sense. But there are many words which are sometimes used in a technical sense and sometimes in a more colloquial or general sense—'guarantee', 'estate' and 'possession' spring to mind as examples. In the context of a statutory provision such as section 209 of the Companies Act 1989 and the Uncertificated Securities Regulations, 'transfer' surely *is* a legal term. As with all statutory legal terms, its meaning is a

[29] Section VIBc *ad fin*.

matter of construction in the context of the relevant provision. Adopting that approach, there seems to be no reason to conclude that the nature of a transfer must be conformed to a pre-existing recognised legal technique—indeed, the presumption may if anything be the reverse, since if that had been the intention the draftsman would surely have used the term intended.

The analysis of a statutory transfer as a novation would have some other strange consequences. A transfer of a paid up share, for example, would involve a cancellation and reissue—an operation which in other contexts requires a court-approved reduction of capital. A transfer of a debt security would involve a fresh issue, valid (it would seem) only if the capacity of the issuer, the authority of its officers and the scope of its procedural authorities were sufficiently wide to encompass the additional volumes of securities involved (which could over time add up to a large multiple of the nominal size of the issue).

For Dr Benjamin, analysis of a transfer as a novation explains why a transferee takes free from any equities of the issuer arising from pre-transfer relations between the issuer and the transferor. If a transfer is not analysed as novation, what is the position in relation to pre-transfer equities of the issuer?

The answer, in the view of the present writer, is that such equities may in some cases be excluded by contract, through the articles or terms of issue; in other cases they may be excluded by estoppel; and there may be cases where they are not excluded at all.

It is less easy than one might perhaps have expected to find examples of exclusion of pre-transfer equities in articles of association. Consider, however, the Table A provisions relating to liens, which provide that the company has a lien on shares which are not fully paid for all amounts owing to it;[30] and that the company may refuse to register a transfer of a share on which it has a lien.[31] It is clear that the company does not have a lien on fully paid shares; and it seems permissible to infer that the right to refuse a transfer precludes the assertion of a lien in respect of amounts owing from the transferor after the transfer has actually been registered.

A clear example of equities being excluded by estoppel is the estoppel arising in favour of a person who has accepted a transfer of shares on the faith of a certificate issued by the company to the transferor, discussed by Dr Benjamin in Section VIBb This estoppel typically arises where shares are transferred under an initial, defective transfer which results in the issue of a certificate and are further transferred by the use of that certificate. In the absence of such a second transfer or of some other change of position in reliance on a representation by the company, the defective transfer is reversible as between the company and the transferee. This may be because

[30] Regulation 8 of the 1985 version of Table A.
[31] Regulation 24 of the 1985 version of Table A.

the transferee is responsible for the defect, but there may also be cases where the company's ability to reverse may be asserted even against an innocent transferee who has not altered his position in reliance on the registration of the transfer.

V. SECURITY OF TRANSFER

In Section VIC Dr Benjamin reviews the different principles originating from the common law, equity and the law merchant relating to the tension between security of title and security of transfer. Her conclusion (with which the present writer agrees) is that:

> the general principles of English law do not deal adequately with the question of third party claims. Security of transfer is reduced, and legal risk is increased, as the outcome of any dispute may depend arbitrarily on the operational detail of the parties' custody arrangements. Generally, this reflects the need for a uniform law of priorities, which does not differentiate between legal and equitable interests.[32]

As Dr Benjamin also notes, a satisfactory solution, in the form of the protection given by the Uncertificated Securities Regulations to properly authenticated dematerialised instructions, was found in the case of dematerialisation of money market instruments. This is not, however, a complete answer to Dr Benjamin's basic point. The relevant provision of the USRs is itself complicated and technical. It also relied on the language of estoppel—in principle a less than ideal formulation of a rule of substantive law.[33] Dr Benjamin's call for a uniform set of rules remains cogent.

[32] Dr Benjamin, Section VID. The present writer has actually had occasion to advise a client whose protection against third party claims of which it had been unaware at the time of transfer depended on whether the client had acquired its interest through a transfer across accounts of the same custodian (with the result that both interests would be equitable and the maxim that 'where the equities are equal, the first in time prevails' would apply) or through a different custodian or nominee which had needed to take a transfer on the register of the issuer (in which case the client might claim to be a bona fide purchaser of a legal interest without notice). It is not easy to explain to clients the relevance of such arcane and historical distinctions.

[33] The draftsman is, of course, in good company, the language of estoppel also being liberally employed in the Bills of Exchange Act 1882—but that Act was a codifying statute, so that the draftsman's hands were tied by the historical development of the relevant rules.

11

Material Adverse Change Clauses After 9/11

RICHARD HOOLEY

I. INTRODUCTION

SINCE THE TRAGIC events of 11 September 2001, material adverse change (MAC) clauses have been in the news. There are plenty of examples.[1] Perhaps the best known is when WPP, the advertising and marketing group, tried to invoke a MAC clause to back out of its bid for Tempus, the media buying group. The Takeover Panel had to consider whether there had been a material adverse change in the prospects of Tempus after the announcement of WPP's offer and, in particular, following the terrorist attacks in New York City and elsewhere in the United States on September 11. The Takeover Panel decided that WPP had failed to demonstrate that there had been a material adverse change in the context of the bid such as to entitle it to invoke the MAC.[2]

What are MAC clauses? In essence, they allow one side or the other the right to pull out of a transaction should unforeseen circumstances prejudice the reason for entering into it. MAC clauses are commonly found in merger and acquisition agreements and in financing agreements with banks, including in banking commitment letters.[3] They are not so common in bond documents. There are two reasons for this.[4] First, because MAC clauses usually

[1] Dynergy, the energy trading company, relied on a MAC clause to get out of its merger with Enron; UBS insisted on MAC language in a financing commitment to support EchoStar's US$25 billion bid for Hughes Electrical: this led EchoStar to seek finance elsewhere (both examples are taken from David Morley, 'Clauses with potent effects', *Financial Times*, 4 March 2002).
[2] *Offer by WPP Group plc for Tempus Group plc* (Takeover Panel Ruling, 6 November 2001).
[3] Commitment letters can constitute binding contracts between a lender and a borrower. The legal character of commitment letters depends on their individual nature: see R Cranston, *Principles of Banking Law* 2nd edn, (Oxford, OUP, 2002), pp 301–303.
[4] PR Wood, *The Law and Practice of International Finance* (London, Sweet & Maxwell, 1980), p 168.

give one party a wide discretion as to the determination of whether a material adverse change has taken place, and trustees for bondholders are generally unwilling to be vested with wide discretions of this sort. Secondly, borrowers are reluctant to place such an uncertain weapon in the hands of a mass of bondholders whom they do not know.

A MAC clause in an acquisition agreement aims to give the buyer the right to walk away from the acquisition before closing if events occur that are detrimental to the target company. The MAC clause fulfils a similar function in a loan agreement, giving the bank the right to suspend or even walk away from its commitment to lend. The absence of a material adverse change is generally a condition to lending under most loan agreements. Moreover, a MAC event of default allows the bank to call a default under a loan if events occur which have, or may have, a prejudicial effect on the borrower's ability to perform its obligations under the financing documents.

There are many variations of the wording of MAC clauses but they are generally wide in scope and vague. This has led to uncertainty and unpredictability as to when such clauses can be relied upon. This is not surprising. In an article published in the *Financial Times* in March 2002, David Morley, managing partner of the global banking practice at Allen & Overy, explained why MAC clauses are so universal but apparently so unpredictable:[5]

> The simple answer is that they are intended to cover risks that people cannot anticipate. In a financing commitment, for example, they may cover an unforeseen change of circumstances affecting the borrower's credit standing or the ability of the lender to spread the risk by bringing in other banks into the deal.
>
> But the real reason most MAC clauses lack predictability is clear: to get deals done. The more explicitly the MAC is defined—by linking it to a specific fall in the FTSE 100, say, or a credit rating downgrade—the more effective and predictable is the clause.
>
> The snag is that insisting on such specific terms can be a deal-breaker. When people buy or sell businesses, or borrowers raise finance, they want to be certain the other side will stick to the deal—specific conditions that weaken this commitment are often negotiated out to get the deal done. What is left tends to follow a basic formula, but with infinite variations that affect the MAC clause's impact.

The human cost of the terrorist attacks on September 11 is beyond measure. The economic cost was also high.[6] In the US, there was already an economic slowdown and this was exacerbated by the terrorist attacks. Global stock markets suffered. Certain sectors were hit hard, namely airlines, hotels and leisure, and insurance. Investors' appetite for risk in the midst of continuing uncertainty has also waned seriously.

[5] David Morley, 'Clauses with potent effects', *Financial Times*, 4 March 2002.
[6] See David Fewtrell, HSBC Investment Bank plc, in *LMA News*, December 2001.

Banks became nervous. In some cases, previous credit risk assessments had become redundant overnight. Banks wanted to know where they stood. The broadly worded MAC clause had got the deal done, but would it now give the bank the protection it needed? Must the bank stand by its commitment to lend if the borrower was now staring financial ruin in the face? What if the bank relied on the MAC clause, refused to lend or declared an event of default, and was later held to have got it wrong? What would be the bank's liability to the borrower? These were some of the questions that the banks asked their lawyers. The MAC clause was put firmly under the spotlight.

This paper has three modest aims. First, to examine the common law rules relating to the frustration of contracts in order to understand the function and operation of the MAC clause better. Secondly, to review some of the common wording employed in MAC clauses in order to get a better idea of when they will operate. In this context it will be important to consider how the courts and other bodies, most notably the Takeover Panel, are likely to interpret and apply such clauses. Thirdly, to consider the potential liability for getting it wrong, in other words, invoking the MAC clause when you are not entitled to do so.

II. FRUSTRATION OF CONTRACTS

Chitty on Contracts states that:[7]

> A contract may be discharged on the ground of frustration when something occurs after the formation of the contract which renders it physically or commercially impossible to fulfil the contract or transforms the obligation to perform into a radically different obligation from that undertaken at the moment of entry into the contract.

In *J Lauritzen AS v Wijsmuller BV (The Super Servant Two)*, Bingham LJ described the essence of the doctrine of frustration through the following five propositions.[8] First, the doctrine has evolved 'to mitigate the rigour of the common law's insistence on literal performance of absolute promises' and its object was 'to give effect to the demands of justice, to achieve a just and reasonable result, to do what is reasonable and fair, as an expedient to escape from injustice where such would result from enforcement of a contract in its literal terms after a significant change in circumstances'. Secondly, frustration operates to 'kill the contract and discharge the parties from further liability under it' and that, therefore, it cannot be 'lightly invoked'

[7] HG Beale *et al*, *Chitty on Contracts* 28th ed, (London, Sweet & Maxwell, 1999), v 1, para 24–001.
[8] [1990] 1 Lloyd's Rep 1 at 8.

but must be kept within 'very narrow limits and ought not to be extended'. Thirdly, frustration brings a contract to an end 'forthwith, without more and automatically'. Fourthly, 'the essence of frustration is that it should not be due to the act or election of the party seeking to rely on it' and it must be some 'outside event or extraneous change of situation'. Fifthly, a frustrating event must take place 'without blame or fault on the side of the party seeking to rely on it'.

A MAC clause operates differently from the doctrine of frustration. First, the MAC clause does not 'kill the contract'. It may relieve one party of his obligation to perform but it does not necessary relieve the other party of his obligations under the agreement. Secondly, the party entitled to rely on the MAC clause will usually have to elect to do so (eg where the bank must declare an event of default). But he may waive this right. Thirdly, the express wording of the MAC clause should be paramount so that if there has been a material adverse change as defined in the clause the party entitled to rely on the clause should be entitled to do so. Save for a particular context that imports its own rules as to when the clause may be invoked, eg where there is a public offer subject to the Takeover Code, there should, in theory, be less scope for the courts to deny a party the right to rely on a clause freely agreed between the contractual parties. However, there are limits to this laissez-faire approach, eg in cases of misrepresentation or unconscionability, and the vaguer the wording of the clause the less likely a court will stand back. In practice this is a real problem, for the MAC clause is often vaguely worded so as to cover unforeseen risks. In any event, there are also certain fetters on how a party may exercise his discretion when the MAC clause is subjective in its wording.[9]

In *The Super Servant Two*, Bingham LJ stressed that the doctrine of frustration cannot be 'lightly invoked' but must be kept within 'very narrow limits and ought not to be extended'.[10] It is not surprising, therefore, that the courts have developed a narrow test for when it can be invoked. There is now general agreement that the appropriate test to apply to determine whether a contract has been frustrated is that of a 'radical change to the obligation'.[11] In *Davis Contractors Ltd v Fareham UDC*, Lord Radcliffe said:[12]

> Frustration occurs whenever the law recognizes that without default of either party a contractual obligation has become incapable of being performed

[9] See section V below.
[10] [1990] 1 Lloyd's Rep 1 at 8, citing *Bank Line Ltd v Arthur Capel & Co* [1919] AC 435, 459; *Davis Contractors Ltd v Fareham UDC* [1956] AC 696, 715, 727; and *Pioneer Shipping Ltd v BTP Tioxide Ltd* [1982] AC 724, 725.
[11] There has been much discussion of the theoretical basis of the doctrine of frustration. Various theories have been advanced, but no practical consequences seem to follow from the debate. See GH Treitel, *Frustration and Force Majeure* (London, Sweet & Maxwell, 1994), pp 578–584; GH Treitel, *The Law of Contract* 10th edn, (London, Sweet & Maxwell, 1999), pp 858–862.
[12] [1956] AC 696 at 729.

> because the circumstances in which performance is called for would render it a thing radically different from that which was undertaken by the contract. *Non haec in foedere veni*. It was not this that I promised to do.

This test has been adopted by the House of Lords in several cases,[13] and in *National Carriers Ltd v Panalpina (Northern) Ltd*, Lord Simon restated the test as follows:[14]

> Frustration of a contract takes place when there supervenes an event (without default of either party and for which the contract makes no sufficient provision) which so significantly changes the nature (not merely the expense or onerousness) of the outstanding contractual rights and/or obligations from what the parties could reasonably have contemplated at the time of its execution that it would be unjust to hold them to the literal sense of its stipulations in the new circumstances; in such case the law declares both parties to be discharged from further performance.

Under this test the court must first construe the terms of the contract in the light of its nature and the relevant surrounding circumstances when it was made. This enables the court to determine the original obligation of the parties. The court must then consider whether there would be a radical change in that obligation if performance were enforced in the circumstances which have subsequently arisen. But a mere rise in cost or expense is not enough. As Lord Radcliffe said in *Davis Contractors*:[15]

> It is not hardship or inconvenience or material loss itself which calls the principle of frustration into play. There must be as well such a change in the significance of the obligation that the thing undertaken would, if performed, be a different thing than that contracted for.

There are several good examples of where the courts have rejected a plea of frustration on the ground that all that has happened is increased hardship or inconvenience. *Davis Contractors Ltd v Fareham UDC* is one such case.[16] In July 1946, a builder entered into a contract with Fareham UDC to build 78 houses for a fixed sum of £94,424. Owing to an unexpected shortage of skilled labour and of certain materials, the contract took 22 months to complete instead of the 8 months expected and cost some £115,000. The House of Lords held the contract had not been frustrated as the builder contended. The mere fact that unforeseen circumstances had

[13] *Tsakiroglou & Co Ltd v Noblee Thorl GmbH* [1962] AC 93, 131; *Pioneer Shipping Ltd v BTP Tioxide Ltd* [1982] AC 724; *Paal Wilson & Co A/S v Partenreederei Hannah Blumenthal* [1983] 1 AC 854, 909, 918.

[14] [1981] AC 675 at 700.

[15] [1956] AC 696 at 729.

[16] [1956] AC 696.

made the contract more onerous to the builder did not discharge the agreement. The thing undertaken was not, when performed, different from that contracted for.

The Suez Canal cases illustrate the same point. In *Tsakiroglou & Co Ltd v Noblee Thorl GmbH*,[17] sellers agreed to sell a quantity of groundnuts to be shipped from Sudan to Hamburg during November or December 1956. On 2 November, the Suez Canal was closed and remained closed for the next 5 months. The price of groundnuts cif Hamburg was clearly calculated on the basis of shipment via the canal, but the contract contained no term to this effect. The sellers refused to perform the contract, claiming that it was frustrated by the closure of the canal. The House of Lords held that there was no frustration since it would be possible to ship the nuts to Hamburg via the Cape of Good Hope. Such a journey would not be commercially or fundamentally different from that by the canal, but merely more expensive. On the other hand, if the goods had been perishable, or if a definite date for delivery (rather than shipment) had been fixed, or if there had been a shortage of shipping to carry goods from Port Sudan to Europe via the longer route, the contract may well have been frustrated.[18]

The courts have remained consistent in their unwillingness to regard increased costs or financial hardship as a ground for frustration of a contract. Increased costs caused by terrorist activities are no exception to this general rule. *Globe Master Management Ltd v Boulous-Gad Ltd*[19] illustrates the point. G agreed to provide B with crew for a ship chartered by B and used to carry Israeli citizens around the Eastern Mediterranean on gambling cruises. Six months after the agreement was signed the Al Aqsa Intifada began with an increased incidence of suicide bombings. B stopped the cruises and laid off the crew. B claimed that the hostilities and dangerous security situation in the area constituted a frustrating event and called on evidence from security experts that the threat of terrorism made an Israeli operation of tourists and cruise ships in the area extremely hazardous and that the passengers could not be protected for an economic price. The Court of Appeal held that there had been no frustrating event. The deterioration in the security situation in the Eastern Mediterranean was not a surprising, let alone radically different, event. There was no indication that underwriters would decline cover for shipping; rather, the evidence relied on focused on the economic cost of providing extra security.

A bank would be unlikely to succeed with a plea of frustration where its own funding costs increased because of the increased cost of funds in the market (in the rare case where the increase could not be passed on to

[17] [1962] AC 93.

[18] HG Beale *et al*, *Chitty on Contracts* 28th edn, (London, Sweet & Maxwell, 1999), v 1, para 24–046 (citing *Ocean Tramp Tankers Corporation v V/O Sovfracht (The Eugenia)* [1964] 2 QB 226, 240, 243).

[19] [2002] EWCA Civ 313.

the borrower) or where the ability of the borrower to repay was threatened because of a downturn in his business or his industry in general. These are risks that the creditor would be expected to carry. However, in *Staffordshire Area Health Authority v South Staffordshire Waterworks Co*,[20] where the defendants agreed in 1929 to provide water 'at all times hereafter' at a fixed price which by 1979 was well out of line with the prevailing price because of inflation, Lord Denning MR expressed the opinion that by reason of 50 years of continuing inflation, a fundamentally different situation had emerged in which the contract had ceased to bind.[21] His reasoning was not accepted by the other members of the Court of Appeal. The orthodox view is that the creditor must bear the risk in any depreciation in the purchasing power of sterling,[22] or the devaluation of a foreign currency in which a debt is expressed.[23] If the creditor does not wish to bear the risk, provision must be made in the contract.[24]

III. FORCE MAJEURE CLAUSES

Problems caused by the narrow ambit of the doctrine of frustration can be avoided by the use of a *force majeure* clause.[25] Such a clause entitles one (or both) of the parties to be excused from performance of the contract, in whole or in part, or entitles him to suspend performance or to claim an extension of time for performance, on the happening of various specified events or events outside his control, even if those events are non-frustrating events. *Force majeure* clauses are frequently inserted into commercial contracts.

Force majeure is not a term of art in English law, although it is well known in continental legal systems.[26] It has, for example, a relatively well defined meaning under French law. A failure of performance by a party to a contract governed by French law will be attributable to *force majeure* if, without any fault on the part of the party seeking to be excused performance, an event occurs that possesses the following characteristics: (i) *irrésistibilité*—the event must render performance of his obligation impossible, and not merely more onerous; (ii) *imprévisibilité*—the event must not be reasonably foreseeable, for he ought then to have taken steps to prevent or avoid it;

[20] [1978] 1 WLR 1387.
[21] At pp 1397–8. However, Lord Denning did not hold that the contract had terminated automatically, reasonable notice was required.
[22] *Wates Ltd v GLC* (1987) 25 BLR 1, 35.
[23] *British Bank of Foreign Trade Ltd v Russian Commercial and Industrial Bank* (1921) 38 TLR 65.
[24] J Beatson, *Anson's Law of Contract* 28th edn, (Oxford, OUP, 2002), p 550.
[25] See GH Treitel, *Frustration and Force Majeure* (London, Sweet & Maxwell, 1994); EMckendrick (ed) *Force Majeure and Frustration of Contract* 2nd edn, (London, LLP, 1995).
[26] AG Guest *et al*, *Benjamin's Sale of Goods* 6th edn, (London, Sweet & Maxwell, 2002), para 8–097.

(iii) *extériorité*—the event must proceed from some external cause, ie, not from a cause within his sphere of responsibility.[27] In other words, French law requires *force majeure* to be 'unforeseeable, insurmountable and irresistable'.[28]

When we turn to English law, it is clear that *force majeure* clauses bring certain advantages to those who rely on them. The first is the obvious advantage that the event specified in the clause can be a non-frustrating event. In other words, unlike the position under French law, a *force majeure* clause can excuse one party from his obligation to perform where performance has become more onerous and not necessarily impossible. Secondly, again unlike the position under French law, the *force majeure* clause can apply under English law to events that are foreseeable.[29] Thirdly, it has been held that a *force majeure* clause does not constitute an exclusion clause.[30] Whether such a clause entirely escapes the clutches of the Unfair Contract Terms Act 1977 is more debatable.[31] However, in the case of a commercial contract it seems unlikely that a clause which merely permits one party to suspend, postpone or cancel performance upon the happening of events beyond his control would be held to be unreasonable in the absence of special circumstances.[32]

Is a MAC clause a *force majeure* clause? As there is no magic in the term *force majeure* under English law, such clauses may assume a variety of forms.[33] Perhaps the best that can be said by way of general description is that a *force majeure* clause makes provision for non-performance or varied performance in circumstances which are beyond the control of the promisor.[34] In substance, this is what a MAC clause does. However, in

[27] This statement of the requirements of *force majeure* under French law is taken from *Benjamin's Sale of Goods* (2nd edn), para 664, as cited by Staughton J in *Navrom v Callitsis Ship Management SA (The Radauti)* [1987] 2 Lloyd's Rep 276, 281–2 (affd [1988] 2 Lloyd's Rep 416).

[28] B Nicholas, 'The Vienna Convention on International Sales Law' (1989) 105 LQR 201 at 235; B Nicholas, *French Law of Contract* 2nd edn, (Oxford, Clarendon Press, 1992), pp 202–3.

[29] See the main text at the reference to n 61 below.

[30] *Fairclough, Dodd & Jones Ltd v JH Vantol Ltd* [1957] 1 WLR 136, 143. Professor Treitel considers *force majeure* clauses to be duty defining clauses rather than exemption clauses: GH Treitel, *Frustration and Force Majeure* (London, Sweet & Maxwell, 1994), para 12–014; GH Treitel, *The Law of Contract* 10th edn, (London, Sweet & Maxwell, 1999), p 839, n 51.

[31] Contrast HG Beale *et al*, *Chitty on Contracts* 28th edn, (London, Sweet & Maxwell, 1999), v 1, para 14–141 (caught by s 3(2)(b) of the Act) with GH Treitel, *Frustration and Force Majeure* (London, Sweet & Maxwell, 1994), para 12–014 (not subject to the Act).

[32] Eg in an exclusive dealing agreement where the supplier is entitled to suspend performance in the event of *force majeure* but the purchaser is not entitled, during the period of suspension, to acquire supplies from another source: see HG Beale *et al*, *Chitty on Contracts* 28th edn, (London, Sweet & Maxwell, 1999), v 1, para 14–141; J Beatson, *Anson's Law of Contract* 28th edn, (Oxford, OUP, 2002), p 194.

[33] A term 'the usual *force majeure* clauses to apply' has been held void for uncertainty: *British Electrical and Associated Industries (Cardiff) Ltd v Patley Pressings Ltd* [1953] 1 WLR 280.

[34] E McKendrick (ed), *Force Majeure and Frustration of Contract* 2nd ed, (London, LLP, 1995), p 10.

merger and acquisition agreements, and in financing agreements with banks, MAC clauses take on a different form to the orthodox *force majeure* clauses which can be found, for example, in the standard terms of many trade associations.[35] In M & A or financing agreements the MAC clause may take the form of a condition of an offer (thereby preventing the performance obligation arising in the first place), or a warranty or representation, or even an event of default on the part of the other contracting party. Moreover, the party relying on the MAC clause may be given particular rights which he would not be expected to have in the case of an orthodox *force majeure* clause. Thus, where a MAC clause triggers an event of default in a loan agreement, the bank is not only relieved of its obligation to advance further sums but it is also entitled to accelerate repayment of the loan. With secured lending the agreement will generally also give the bank a power to appoint a receiver or an administrator.[36] These differences of form are significant and must be kept in mind.

IV. SPECIMEN MAC CLAUSES

MAC clauses come in all shapes and sizes. In each case it is important to consider the actual wording of the particular clause and the terms of the agreement as a whole. A typical example of a MAC clause for private acquisitions in the UK runs as follows:[37]

> There shall not have occurred any material adverse change in the business, operations, assets, position (financial, trading or otherwise), profits [or prospects] of the Target Group taken as a whole [or any event or circumstances that may result in such a material adverse change]
>
> [excluding, in any such case, any event, circumstance or change resulting from:
>
> a. changes in stock markets, interest rates, exchange rates, commodity prices or other general economic conditions, or
> b. changes in conditions generally affecting the [industry];

[35] Eg, GAFTA Contract 100, clauses 21 and 22, the clauses generally relied upon in the soya bean embargo cases: see M Bridge, 'The 1973 Mississippi Floods: 'Force Majeure' and Export Prohibitions' in E McKendrick (ed), *Force Majeure and Frustration of Contract* 2nd edn, (London, LLP, 1995), ch 15.

[36] R Cranston, *Principles of Banking Law* 2nd edn, (Oxford, OUP, 2002), p 322. See the Enterprise Act 2002.

[37] From www.practicallaw.com (the bracketed words may be subject to negotiation). The MAC clause seeks to allow the buyer to withdraw from the purchase if there is a material adverse change of the business, assets or profits of the target between agreeing to purchase the company and completion of the sale. In other words, it is only relevant where there is a gap between signing and closing.

c. changes in laws, regulations or accounting practices;
d. matters disclosed [in the Disclosure Letter or the Acquisition Documents]; or
e. this transaction or the change in control resulting from this transaction]

> [except to the extent that the matters in paragraphs (a) to (c) have an impact on the Target Group that is disproportionate to the effect on other [similar] companies operating in the [industry].]

A MAC event of default clause in a loan agreement may be worded as follows:[38]

> event of default if any change occurs, financial or otherwise, which affects the borrower in a manner which in the opinion of the bank is materially adverse.

Alternatively, the clause could read as follows:[39]

> event of default if any event or series of events occurs which, in the opinion of the Majority Banks, might have a material and adverse effect on the financial condition of a member of the Group or on the ability of the Borrower to comply with its obligations under the finance documents.

V. SUBJECTIVITY

The MAC clause may be phrased so that the triggering event or change is determined subjectively by the bank. Alternatively, the clause may employ an objective test. The bank will prefer the former, the borrower the latter (arguing that a subjective test turns a term loan into an on-demand facility). There may also be occasions where third persons are required by the clause to declare that the material change has arisen.

The borrower will be particularly worried where the clause is subjective. As Buchheit states:[40]

> The banker who exhumes the loan agreement from his files in order to verify the existence of a MAC event of default is, *ex hypothesi*, the banker who is

[38] From PR Wood, *International Loans, Bonds and Securities Regulation* (London, Sweet & Maxwell, 1995), para 3–48.
[39] *Ibid*, Appendix, Pt I.
[40] LC Buchheit, 'How to negotiate the material adverse change clause' [1994] IFLR (March) 31 at 32; LC Buchheit, *How to Negotiate Eurocurrency Loan Agreements* 2nd edn, (London, Euromoney Publications, 2000), p 111.

> feeling some rabbit in his blood. The circumstances under which the lenders are likely to be asked to decide whether a material adverse change has occurred are precisely the circumstances in which the lenders will be emotionally predisposed to find such a change.

Some protection may be given to the borrower where the declaration of a change is to be made by a super-majority of banks involved in a syndicated loan.

However, it would be wrong to think that a bank seeking to rely on a MAC event of default, for example, has an unfettered discretion to declare that there has been a material adverse change even where the test is a subjective one. In some cases the MAC clause expressly states that an event of default arises when in the bank's 'reasonable' opinion there has been a material adverse change. In any event the word is likely to be read into the clause whether it expressly appears or not.

In the United States defaulting borrowers have sometimes claimed damages for a lender's alleged failure to meet its implied obligation of 'good faith' in dealing with a customer.[41] Such claims are typically made on termination of a line of credit, upon acceleration of payment under a note, or at foreclosure on collateral.[42] Bankers have been held liable even when exercising an express contractual power under a loan agreement. The duty of good faith can arise under common law[43] or commercial statute.[44]

English law knows no general doctrine of good faith.[45] By contrast, English law has 'developed piecemeal solutions in response to demonstrated problems of unfairness'.[46] The application of concepts such as election, estoppel, waiver, relief against forfeiture etc are examples of this piecemeal approach.

Under English law the terms of the loan facility are paramount.[47] The courts will construe those terms to give effect to the intention of the parties. But occasionally the courts will imply a term into the agreement to control

[41] JJ Norton, 'Lender Liability in the United States', ch 12 in W Blair (ed), *Banks, Liability and Risk* 3rd edn, (London, LLP, 2001). See also DR Fischel, 'The Economics of Lender Liability' (1989) 99 *Yale Law Journal* 131.
[42] See eg, *Reid v Key Bank*, 821 F 2d 9 (1st Cir 1987); *KMC v Irving Trust*, 757 F 2d 752 (6th Cir 1985); *Martin Speciality Vehicles v Bank of Boston*, 87 *Bankruptcy Reporter* 752 (Bankruptcy Ct D Mass 1988), revsd on other grds, 97 *Bankruptcy Reporter* 721 (D Mass 1989); *Duffield* v *First Interstate Bank of Denver*, 13 F 3d 1403 (10th Cir 1993), cert denied, 13 June 1994.
[43] The duty usually arises under state law rather than federal law. Restatement (Second) of Contracts, s 205 provides: 'Every contract imposes upon each party a duty of good faith and fair dealing in its performance and enforcement'.
[44] See Uniform Commercial Code, s 1–203: 'Every contract or duty within this Act imposes an obligation of good faith in its performance or enforcement'.
[45] *Walford v Miles* [1992] 2 AC 128, 138.
[46] *Interfoto Picture Library Ltd v Stiletto Visual Programmes Ltd* [1988] 1 All ER 348, 353, per Bingham LJ.
[47] There is some statutory control in the case of loans made to individuals which do not exceed £25,000 (see the Consumer Credit Act 1974), and in the case of unfair terms in consumer loan agreements (see the Unfair Terms in Consumer Contracts Regulations 1999, SI 1999 No 2083). The Unfair Contract Terms Act 1977 catches unreasonable exemption and limitation clauses

the risk of unconscionable behaviour by one party or the other. The recent Court of Appeal decision in *Paragon Finance plc v Staunton*,[48] provides a good example of this. There the lender's discretionary power to vary the interest rate in a residential mortgage was controlled by an implied term that the lender would not act dishonestly, for an improper purpose, capriciously, arbitrarily or unreasonably. The Court of Appeal held that the lender was not to exercise his discretion in a way no reasonable lender, acting reasonably, would do (which was declared to be analogous to unreasonableness in the *Wednesbury* sense). Their Lordships added that it was unlikely that a lender who was acting unreasonably would not also be acting either dishonestly, for an improper purpose, capriciously or arbitrarily.

VI. MATERIALITY

The adverse change must be 'material'. What does this mean? The dictionary defines material to mean 'important, relevant and essential'. The case law seems to provide a similar definition. In *Tesco Stores Ltd v Secretary of State for the Environment*,[49] for example, the House of Lords held that for the purposes of s 70(2) of the Town and Country Planning Act 1990, which requires the Secretary of State to have regard to 'material considerations' when dealing with applications for planning permission, a 'material' consideration meant a relevant consideration. In *DB Rare Books Ltd v Antiqbooks (a limited partnership)*,[50] the Court of Appeal held that a 'material' breach arose under the terms of a partnership agreement where the breach was of 'serious or substantial import' or 'important'. However, caution must be exercised when seeking to define a term used in one statute or agreement for the purpose of defining the same term in another. The warning is clear: 'No useful progress can be made … by inquiring what meaning the courts have given [to a particular word] in reported cases, for they draw their meaning entirely from the purpose for which and the context in which they are used'.[51]

Some useful guidance as to materiality in this context can be found in those cases where the courts have held that a material adverse change has occurred. *National Westminster Bank Ltd v Halesowen Presswork Assemblies Ltd* is such a case.[52] In April 1968 the claimant's current account

in standard form loan agreements and/or loan agreements made with consumers. See generally, EP Ellinger, E Lomnicka and RJA Hooley, *Modern Banking Law* 3rd edn, (Oxford, OUP, 2002), ch 16.

[48] [2001] EWCA Civ 1466, [2001] 2 All ER (Comm) 1025.
[49] [1995] 2 All ER 636.
[50] [1995] 2 BCLC 306.
[51] *Southern Water Authority v Nature Conservatory Council* [1992] 1 WLR 775 at 781, per Lord Mustill, quoting *Madrassa Anjuman Isamia v Municipal Council of Johannesburg* [1922] 1 AC 500 at 504, per Viscount Cave.
[52] [1972] AC 785.

showed a substantial debit balance and the defendant bank agreed to open a no 2 account for the claimant's trading operations. The bank agreed that, in the absence of a material change of circumstances, account no 1 would remain frozen for a period of four months. On 20 May the claimants convened a meeting of their creditors. The bank, which received notice of the meeting, resolved to leave the April arrangement in place. On 12 June the claimants passed a resolution to wind up voluntarily. On 19 June the bank informed the liquidator that it had determined to set off the credit balance in account no 2 against the debit balance in account no 1. The liquidator objected, as in his view the bank was bound by the April arrangement. The House of Lords held that the defendant bank had the right to combine the accounts both in view of the nature of the arrangements made in April and under s 31 of the Bankruptcy Act 1914 (now s 323 of the Insolvency Act 1986). The resolution to wind up constituted a material change of circumstances, and the bank's right to combine revived forthwith.

In *Levison v Farin*,[53] the claimants were the shareholders of L Co, a fashion company, which marketed clothes designed by L, a designer of high reputation. L became ill and could not work, and L Co began to run at a loss. The claimants negotiated to sell L Co to the defendants. They disclosed the illness of L, that no new collection had been designed, that L Co was running at a loss and that it was in a traded down state which they could not specify because they did not know how much money was being lost. The defendants agreed to buy L Co's assets for £44,000 and its goodwill for £10,000. The agreement contained a warranty that 'save as disclosed as between balance sheet date and completion date there will have been no material adverse change in the overall net assets of the company allowing for normal trade fluctuations'. When the defendants took over they found an adverse change of £8,600. They refused to pay the full purchase price. Gibson J held that the defendants were entitled (on their counterclaim) to damages for breach of the warranty. The reduction was not a normal trade fluctuation but a material adverse change. The judge considered a 20 per cent drop in net asset value in about four months, between the date of the last financial accounts and the completion date of the transaction, to be 'material'.

Levison v Farin is a particularly interesting decision. The claimants had argued that even if there had been a material adverse change there was no breach of warranty because there had been disclosure to the defendants. Gibson J rejected this argument. He held that disclosure of L's illness and the trading down of the company was a general disclosure of the causes of probable future losses and was not disclosure of a quantified reduction in the net asset value or the actual rate of the continuing losses.

[53] [1978] 2 All ER 1149.

This raises the question of whether a MAC clause can apply to an event that could have been foreseen or guarded against. The decision of Gibson J in *Levison v Farin* suggests that it can.[54] A useful comparison can be made with the common law of frustration. In many of the cases reference is made to the occurrence of an 'unforeseen' or 'unexpected' or 'uncontemplated' event. It may be thus argued that where the parties have assumed the risk of an event occurring which was present in their minds at the time the contract was made, the contract cannot be frustrated when the event actually occurs.[55] On the other hand, there are also cases where it was said that a contract could be frustrated by foreseen or foreseeable events.[56] The position is not entirely clear. Professor Treitel accepts that a contract cannot be frustrated by foreseen or foreseeable event but has argued for a high test of foreseeability: 'To support the inference of risk-assumption, the event must be one which any person of ordinary intelligence would regard as likely to occur. Moreover, the event or its consequences must be foreseeable in some detail'.[57] Professor Beatson submits that it is a question of construction of the contract whether it was intended to continue to bind in that event, or whether, in the absence of any express provision, the issue has to be left open, so as to allow the incidence of the risk to be determined by the law of frustration.[58]

We have already seen that under French law *force majeure* must be 'unforeseeable, insurmountable and irresistible'.[59] It may be argued that a party should not be entitled to relief under a *force majeure* clause in respect of an event if, when he entered into the contract, he could reasonably have been expected to take the risk of the event into account and make contingency plans to deal with it.[60] However, English law does not adopt this approach.[61] In *Navrom v Callitsis Ship Management SA (The Radauti)*,[62] the *force majeure* clause in the charterparty referred to 'hindrances … delaying … the discharge

[54] But there are US cases suggesting otherwise: *Re JC's East*, 1995 US Dist LEXIS 13607 (SDNY 1995); *Sinclair Broadcast Group v Bank of Montreal*, 94 Civ 4677 (SDNY 1995). See RM Gray, 'Does the crisis bring default under MAC clauses?' [1998] IFLR (April) 17 at 18.

[55] See eg, *Krell v Henry* [1903] 2 KB 740, 752; *Tamplin SS Co Ltd v Anglo-Mexican Petroleum Co* [1916] 2 AC 397, 424; *Bank Line Ltd v Arthur Capel & Co* [1919] AC 435, 458; *Davis Contractors Ltd v Fareham UDC* [1956] AC 696, 731.

[56] See eg, *WJ Tatem Ltd v Gamboa* [1939] 1 KB 132, 135; *Ocean Tramp Tankers Corporation v V/O Sovfracht (The Eugenia)* [1964] 2 QB 226, 239.

[57] GH Treitel, *Law of Contract* 10th edn, (London, Sweet & Maxwell, 1999), p 841; GH Treitel, *Frustration and Force Majeure* (London, Sweet & Maxwell, 1994), ch 13.

[58] J Beatson, *Anson's Law of Contract* 28th edn, (Oxford, OUP, 2002), p 548.

[59] Section III above.

[60] The argument is set out, but rejected, in E McKendrick (ed), *Force Majeure and Frustration of Contract* 2nd edn, (London, LLP, 1995), p 79.

[61] Cf Art 79 of the Vienna Convention on International Sales; contrast also Council Directive 90/314 on package travel, package holidays and package tours ([1993] OJ L158/59), and the Package Travel, Package Holidays and Package Tours Regulations 1992, SI 1992 No 3288, regs 13(3)(b), 15(2)(c)(i).

[62] [1988] 2 Lloyd's Rep 416.

of the cargo'. Congestion at Tripoli caused a delay with discharge. The owners argued that because congestion was common at Tripoli, this fell outside the clause. Lloyd LJ held that the foreseeability of congestion was irrelevant:[63] 'The phrase "hindrances ... delaying ... the discharge of the cargo" should be given its ordinary meaning, even though on the facts some degree of hindrance was inevitable. The foreseeability of the congestion does not justify attaching an unusual or restricted meaning to the word "hindrances"'.

At first instance Staughton J had examined the requirement under French law that one of the necessary elements of *force majeure* is *imprévisibilité*, in other words, that the event must not be reasonably foreseeable. Staughton J said that insofar as the expression '*force majeure*' even had a general meaning in English Law, he doubted whether it necessarily conveyed this element of *imprévisibilité*:[64] 'Some wars may be foreseen, some strikes and some abnormal tempests or storms. I would suggest that it is more a question of causation; whether the incidence of a particular peril which could have been foreseen can really be said to have caused one party's failure to perform'.

It is submitted that the conclusion to be drawn from this review of the law of frustration and *force majeure* is that a suitably worded MAC clauses could extend cover to foreseeable events. It would be a matter of construction of the wording of the particular clause. General and/or sectoral economic decline that was known to, or should have been foreseen by, the party relying on the clause when he entered into the contract is unlikely to be held to constitute a material adverse change unless the wording of the clause is particularly clear on the point.[65] How common such broadly worded clauses are likely to be in practice is a different matter as the commercial purpose of a MAC clause is usually to deal with events that are unpredictable and unforeseen.[66]

There can be no legal certainty as to whether any given set of circumstances will be determined by a court to constitute a material adverse change. Ultimately the question of materiality will be determined by the court on an objective basis, and evidence would be heard from experienced bankers as to the real materiality in this sort of situation.[67] The more broadly worded the MAC clause the greater the uncertainty. This has led

[63] *Ibid*, 420.
[64] [1987] 2 Lloyd's Rep 276, 282.
[65] The Takeover Panel took a narrow view on the point in its ruling in the WPP/Tempus case (see the main text at the reference to fn 82 below). See also the Delaware Court of Chancery Division in *IBP Inc Shareholders Litigation v Tyson Foods Inc*, No CIV A18373.2001 WL (Westlaw) 675330, Del Ch 18 June 2001.
[66] See Section I above.
[67] R Youard, 'Default in International Loan Agreements I and II' [1986] *JBL* 276 and 376 at 381. An analogy can be drawn with the position in insurance law where there is an allegation that the insured has not disclosed a material fact to the insurer. A fact is material for this purpose if it is one that would influence the judgment of a prudent insurer in deciding whether to undertake the risk and, if so, at what premium: *Lambert v Cooperative Insurance Society* [1975] 2 Lloyd's Rep 485, CA; *Pan Atlantic Insurance Co Ltd v Pine Top Insurance Co Ltd* [1995] 1 AC 501, HL.

some draftsmen to formulate more narrowly worded clauses which provide their own definition of what is material. A MAC clause could be defined, for example, by reference to a certain percentage fall in profits or some defined external economic indicator, eg a drop in the borrower's credit rating. But we have already noted David Morley's observation that insisting on such specific terms can be a deal breaker.[68] There are other dangers. If the clause is too narrow it may exclude important types of changes to the borrower's business.[69] *Re TR Technology Investment Trust plc*[70] is an example of a case where problems were created by excessive definition. There a share purchase was financed by a syndicated loan. Subsequently, the borrower's assets, the purchase of which was financed by the loan, were frozen and an issue arose as to whether this constituted a material adverse change under the loan agreement. The MAC clause in question required that 'the 'adverse change' … be judged by reference to the company's financial statements'. But as the company was newly incorporated, no financial statements had been prepared at the time the assets were frozen. It was not possible to say, following the definition contained in the MAC clause, whether there had been such an adverse change.[71]

VII. WHAT MUST THE CHANGE AFFECT?

This depends on the wording of the MAC clause. The MAC clause may refer to adverse changes (i) in the market; (ii) in the financial conditions of the borrower/target (and its group); or (iii) in national or international financial, economic or political conditions. In fact, the MAC clause may refer to all three.

Bank commitment letters often use a market MAC clause which typically falls away on signing the facility agreement. Immediately following the events of September 11 there were attempts by some banks to amend the terms of commitment letters already sent out to prospective borrowers prior to that date to ensure that any market disruption or change caused by those events was caught by the MAC clause.[72]

Where a bank is providing finance for acquisition purposes the bank may be prepared to ride on the back of the business MAC clause contained in the acquisition documentation. The wording of the business MAC in the

[68] See the main text at the reference to fn 5 above.
[69] See E Kontor and J Day, 'Corporate Lending in an Intangible Economy: Potential Solutions' [2002] *JIBL* 174 at 178.
[70] (1988) 4 BCC 244, Hoffmann J.
[71] *Ibid*, at 264.
[72] In other cases, banks simply relied on the market flex clause. The language of a typical market flex provision is very wide. For example (taken from *LMA News*, December 2001, p 7):

'If the Arranger at any time determines a change is advisable to achieve a successful syndication of the Facility, the Arranger reserves the right, after consulting with the Borrower acquisition documentation could be quite specific, eg X per cent drop in revenues; X per cent drop in equity value. This reduces the level of uncertainty. Ideally the bank would look to have the protection of specific business MAC provisions coupled with the more general and sweeping business MAC language.

The MAC clause is very common in sovereign loans. A version of the clause is standard in the loan agreements of the World Bank and regional development banks.[73] The MAC clause used for sovereign borrowers often refers to changes in the sovereign's economic, financial or political conditions, although some sovereign borrowers find the reference to political conditions offensive in this context.[74]

Two common examples of the MAC clause in a loan agreement are the MAC representation and the MAC event of default. With a MAC representation the borrower represents and warrants that there has been no material adverse change in its business, assets, operations, prospects or condition, financial or otherwise since a specified date. By referring to factors specific to the borrower, it can test to see whether any of those factors have changed in a materially adverse way since the specified date (usually the date of the borrower's latest audited financial statements delivered before signing the financial documentation). As the loan agreement usually requires, as a condition precedent to borrowing, that representations made by the borrower remain true, the borrower will not be entitled to borrow if a material adverse change has occurred.[75] This suspends the bank's obligation to advance the loan.

A MAC event of default usually refers to any event or circumstance that could reasonably be expected to materially and adversely affect the borrower's ability to perform its obligations under the financing documents. In the face of an event of default the bank has the option of terminating the lending commitment and accelerating maturity of the loan. Unlike the MAC representation, the MAC event of default does not usually focus simply on factors specific to the borrower. However, the MAC event of default usually requires that the adverse event or circumstance does affect the ability of the borrower to perform its obligations under the financing documents. But the bank should be careful not to limit such tests to the borrower's ability to pay as this could amount to proving the borrower is insolvent.

for a period not exceeding 5 Business Days to alter the structure, terms and pricing of the Facility (but not the actual amount thereof)'.

[73] PR Wood, *International Loans, Bonds and Securities Regulation* (London, Sweet & Maxwell, 1995), para 3–49.
[74] LC Buchheit, *How to Negotiate Eurocurrency Loan Agreements* 2nd edn, (London, Euromoney Publications, 2000), p 112.
[75] There are examples of highly rated borrowers succeeding in having a day 1 only, non-repeating MAC representation.

The trigger contained within the MAC clause will vary from clause to clause and is often a matter for negotiation. The point is covered by David Morley in his article in the *Financial Times* in the following terms:[76]

> Small differences in wording can have large effects on the magnitude of the MAC clause. Some are activated if the change of circumstances 'might' or 'may' have a material adverse effect. A 1 per cent chance is enough. At the other end of the scale, 'will' means 100 per cent certainty that there will be a material adverse effect, which is hard to prove. A common compromise is 'likely to', which is usually interpreted as meaning a better than evens chance.

VIII. WPP/TEMPUS

The question of whether there had been a material adverse change arose in the context of WPP's bid for the Tempus Group. In July 2001, Havas Advertising made a bid for Tempus. WPP, which already owned 22 per cent of Tempus, countered the Havas bid with a higher bid of its own. WPP announced its offer for Tempus on 20 August, valuing Tempus at approximately £434 million. WPP's offer document was posted on 10 September and WPP bought a further 3 per cent of Tempus in the open market on 17 September. Havas allowed its offer to lapse on 24 September by virtue of not reaching the 90 per cent acceptance threshold, citing the deteriorating market conditions and the events in the US on September 11 as reasons for not submitting a higher bid. By contrast, WPP's offer became unconditional as to acceptances on 2 October after having received acceptances for, or purchased during the offer period, 94 per cent of the Tempus shares to which its offer related.

On 10 October, WPP announced that it was seeking the consent of the Takeover Panel Executive to invoke a condition of its offer and pull out of the deal. The offer had been made conditional on 'no adverse change or deterioration having occurred in the business, assets, financial or trading position or profits or prospects of any member of the wider Tempus Group'. WPP claimed that there had been a material adverse change in Tempus's prospects after the announcement of its offer and, in particular, following the attacks in the US on September 11. On 25 October, the Panel Executive ruled that WPP should not be permitted to invoke the condition. WPP appealed. On 1 November, the Takeover Panel dismissed the appeal.[77]

In reaching its decision the Takeover Panel considered Note 2 to Rule 13 of the Takeover Code, which provides that the circumstances that give rise to the right to invoke a condition must be 'of material significance to the

[76] David Morley, 'Clauses with potent effects', *Financial Times*, 4 March 2002.
[77] The Takeover Panel published its reasons on 6 November 2001: *Offer by WPP Group plc for Tempus Group plc*, 6 November 2001.

offeror in the context of the offer'. The only conditions that the Note does not apply to are the acceptance condition and the typical EC and UK antitrust conditions. The Panel held that:[78]

> meeting [the material significance] test requires an adverse change of very considerable significance striking at the heart of the purpose of the transaction in question, analogous ... to something that would justify frustration of a legal contract.

We have seen that the test for common law frustration—that the thing is radically different from that which was undertaken by the contract—is exceptionally high.[79] Financial loss in performing the contract would not be a frustrating event. This raises the question whether circumstances could ever arise which the Panel would consider sufficiently significant to allow reliance on a typical general MAC condition. Following WPP/Tempus, specific MAC conditions probably stand a greater chance of surviving the 'material significance' test, eg in the takeover of Foseco by Burmah Castrol in 1990 a specific condition was included which related to the price of oil.[80]

Several other points come out of the WPP/Tempus decision.[81] First, it is clear from the Panel's decision that an offeror cannot rely on a general and/or sectoral economic decline that was known to, or should have been foreseen by, it in advance of making its offer.[82] This presents the offeror seeking to rely on the MAC in volatile market conditions with the difficulty of having to distinguish between the effects of foreseeable and unforeseeable events. Secondly, in order to show a material adverse change in an offeree company's prospects, a change in the longer-term prospects (ie beyond the current financial year) of the offeree company must be shown.[83] A similar approach was recently taken by a US court when it compelled Tyson, the US poultry producer, to complete its $3.2 billion acquisition of IBP, the beef-processing business, when Tyson tried to back out of the deal by relying on a MAC clause.[84] But the longer-term effect of particular events will always be uncertain and open to different interpretations. Thirdly, WPP's purchase of 3 per cent of Tempus shares on 17 September was clearly held against it by the Panel. WPP claimed that the purchase had been made for tactical reasons, at a time when the bid from Havas was still

[78] *Ibid*, para 16.
[79] See Section II above.
[80] See K Birkett, 'Untying the knot: material adverse change clauses' [2002] PLC 17 at 23.
[81] For detailed analysis, see Allen & Overy Bulletin, *Material Adverse Change Conditions in Public Takeover Offers Post WPP/Tempus* (London, November 2001).
[82] *Offer by WPP Group plc for Tempus Group plc* (Takeover Panel Ruling, 6 November 2001), para 18.
[83] *Ibid*, paras 31 and 34, although the Panel added the rider that this requirement applied 'at least in the present type of case'.
[84] *IBP Inc Shareholders Litigation v Tyson Foods Inc*, No CIV A18373.2001 WL (Westlaw) 675330, Del Ch 18 June 2001.

on the table, in order to achieve a blocking 25 per cent minority shareholding. However, the Panel considered that the purchase sent a signal to the market and to Tempus shareholders that, notwithstanding the events of September 11, WPP did not then consider that the MAC condition applied and intended to proceed with its offer.[85]

What seems clear is that following the WPP/Tempus decision bidders may have to rely on other grounds to get out of their bid. The acceptance condition usually provides such an escape route.[86] For a bid to succeed, acceptance in respect of the requisite number of shares specified in the acceptance condition must be achieved within 60 days of the posting of the offer document, failing which the offer must lapse, unless the Panel consents to it remaining open.[87] The condition is usually set at 90 per cent of the shares to which the offer relates. WPP's offer for Tempus was unusual in that the 90 per cent acceptance level was reached on the first closing date, ie, 21 days after the offer document was posted. More often, the 90 per cent acceptance level is not met at the first closing date, leaving the bidder with an escape route.

IX. A WIDER APPLICATION OF WPP/TEMPUS?

The Takeover Panel's narrow interpretation of the MAC condition in WPP's offer must be read in context. This was an offer by one public company for the shares of another. The announcement of the offer would undoubtedly have had an effect on the share price of Tempus. Withdrawal of that offer would have contributed to the creation of a false market in Tempus shares. It is submitted that the Panel's concern to avoid the creation of a false market explains why it imposed such a high threshold for invocation of the MAC condition. The Panel's willingness to refer to, and take account of, its own statement made in 1974 bears this out. The 1974 Statement was issued because of concern over offerors failing to post offers which had been announced. It led to the introduction of Rule 2.7 of the Code. Rule 2.7 provides that where there has been an announcement of a firm intention to make an offer, the offeror must, except with the consent of the Panel, proceed with the offer unless the posting of the offer is subject to the prior fulfilment of a specific condition and that condition had not been met. In a statement issued at a time of major market decline in 1974, the Panel said:[88]

> The Panel considers that a change in economic, industrial or political circumstances will not normally justify the withdrawal of an announced offer. To justify

[85] *Ibid*, para 34.
[86] As does a provision that the offer may fail if it is not approved by the offeror's shareholders, if this is also a condition of the bid.
[87] Rule 31.6, Takeover Code.
[88] Panel Statement 1974/2, January 1974. See also Panel Statement 1974/4, March 1974.

> unilateral withdrawal, the Panel would normally require some circumstances of an entirely exceptional nature and amounting to something of the kind which would frustrate a legal contract. It must be remembered that the terms and timing of an announcement of intention to offer and of the posting of offer documents are, subject to the Code, entirely in the hands of the offeror. It is therefore right that an offeror shall accept the risk of a change of circumstances in the intervening period. Once an offer is announced, the market in the shares of the offeree company is likely to be, at least to some extent, supported by the price at which the offer has been fixed. It follows that withdrawal would contribute to the market having been a false one.

It is submitted, therefore, that there were clear policy reasons which explain why the Takeover Panel adopted such a narrow view of WPP's entitlement to rely on the MAC condition contained in the offer document. Similar policy reasons would not be present in a different context, where a court would be much more likely to take a less restrictive approach to the invocation of a MAC clause. Some support for this view can be found in the recent decision of Lloyd J in *Re Perusahaan Perseroan (Persero) PT Persuahaan Penerbangan Garuda Indonesia*.[89] Here, a creditor made an application to adjourn a hearing to approve a scheme of arrangement affecting an Indonesian company which operated the Indonesian national airline. Proceedings in relation to the scheme of arrangement were well under way before 11 September 2001. A meeting of the company's creditors had been held on September 17 in Singapore. At that meeting a number of questions were asked about the implications of the events of September 11. It was made clear, in answer to those questions, that the directors of the company did not intend to recast the company's financial forecast which had been issued as part of their explanatory statement on 1 August. This led to an application being made for adjournment of the scheme of arrangement hearing on the ground that more time was required to obtain information to assess whether the events of September 11 had resulted in a material adverse change in the financial position of the airline. The applicant creditor identified a number of matters which may have had adverse effects on the financial performance of the company, including reduced demand for air travel in general and increased volatility in international aviation fuel prices, and uncertainty in the financial markets at large. The creditor's case was that more time was required to collect further information, especially as to the availability of insurance for airlines in general following the events of September 11, and the impact of increased insurance premiums on this airline in particular.

When the application came before Lloyd J, Leading Counsel for the company accepted the possibility that 'the effects of the reduction of insurance cover or other consequences of the events of September 11 might lead to

[89] Unreported, 3 October 2001 (Chancery Division).

material adverse change in the financial position of the company'.[90] Moreover, Lloyd J expressly stated that the impact of additional insurance premiums was 'a matter which is capable of leading to a material adverse change'.[91] The position remained uncertain. However, Lloyd J was not prepared to adjourn the proceedings because the airline's directors had given an undertaking that the scheme of arrangement would only go ahead if the directors could certify, by the time of the scheme's implementation, that there had been no material adverse change. The Court of Appeal later upheld Lloyd J's decision.[92] The plan for the directors to certify that there had been no material adverse change in the airline's financial position had the support of the vast majority of the scheme creditors and so the court would be slow to say that it was unsatisfactory.[93]

This case also illustrates another point of major practical significance. The applicant required more time to acquire information as to the impact of the events of September 11 on the airline. Most of this information was held by the airline itself. In other words, there are problems of asymmetric information. This may create a real difficulty were a bank seeks to declare an event of default on the basis of a material adverse change in the financial position of the borrower. The information needed to establish that there has in fact been such a change may be held by the borrower and not by the bank. The danger is that the bank may declare an event of default only later to discover that this was not the case. Guesswork can be dangerous. The bank may find itself exposed to a large claim for damages from the borrower if it gets it wrong. Information acquired by the bank after it has declared an event of default may confirm that there had been a material adverse change. Will the bank be able to rely on this subsequently acquired information to justify its earlier decision? There is case law in a different area of contract law which suggest that it may.[94] A party may give a wrong or inadequate reason for refusing to go on with a contract but later rely on a breach of contract committed at the time by the other party even though the breach only came to his knowledge subsequently.[95] Nevertheless, where the MAC clause refers to a change that affects the borrower in a manner

[90] Transcript, at p 6.
[91] Transcript, at p 9.
[92] [2001] EWCA Civ 1696.
[93] *Ibid*, at para 62.
[94] But contrast LC Buchheit, 'How to negotiate the material adverse change clause' [1994] IFLR (March) 31 at 32, and LC Buchheit, *How to Negotiate Eurocurrency Loan Agreements* 2nd edn, (London, Euromoney Publications, 2000), p 110: 'woe unto the lender who decides to accelerate first and verify the existence of the MAC default later' (citing *Windsor Shirt Co v New Jersey National Bank*, 793 F Supp 589, 604, n 17 (ED Pa 1992)).
[95] *Taylor v Oakes Roncoroni & Co* (1922) 127 LT 267, 269. Subject to any defence of estoppel that might be available to the breaching party where he relies on the wrong or inadequate reason given and fails to remedy the actual breach: see *Panchaud Frères SA v Etablissements General Grain Co* [1970] 1 Lloyd's Rep 53, 57–8.

which *in the opinion of the bank* is materially adverse, the bank must honestly and reasonably hold that opinion at the time it makes the declaration of default.[96] It cannot seek to bolster an opinion that was not honestly or reasonably held *ex post facto*.

X. INCORRECT INVOCATION OF THE MAC CLAUSE

Since 9/11 there has been marked concern that banks may be more willing to invoke MAC clauses than was previously the case. This is illustrated by the recent attitude of rating agencies to standby or back up lines of credit that contain MAC clauses. Prior to this, rating agencies accepted that standby or back up lines of credit to support a commercial paper (CP) programme could contain a MAC clause. Now rating agencies are adopting a much more hostile attitude to the MAC clause as they are concerned that it will be triggered to block the availability of funds at the very time access to the back up line is needed. Indeed, some borrowers have used this recent change in attitude of the rating agencies to argue against the inclusion of the MAC clause in the first place. However, there is, as yet, no evidence that the mere existence of a general MAC clause has caused the rating agencies to disregard the availability of a back up line of credit for the purposes of rating the underlying CP programme.

Yet it would be a foolish bank that relied too heavily on a MAC clause when pulling the plug on a borrower. The penalty for getting it wrong could be severe. Buchheit gives a stark warning:[97]

> In this era of lender liability concerns, it is a brave banker who will rely exclusively on the MAC clause as the basis for accelerating a credit. Most commentators agree that in the absence of some other objective event of default, the change affecting the borrower would have to be 'cataclysmic' before a lender could be certain that the materiality test was satisfied.

The bank may rely on the MAC clause to justify withholding funds from the borrower where the absence of a material adverse change is a condition precedent to lending. Alternatively, the bank may rely on the MAC clause to declare an event of default, terminate its commitment to make further advances and accelerate repayment of the loan. In each case, should the bank get it wrong, and find itself in breach of the loan agreement, it will be liable to the borrower for such damages as are necessary to put the borrower in the position that it would have been in if the advance had been

[96] See the main text at the reference to n 48 above.

[97] LC Buchheit, 'How to negotiate the material adverse change clause' [1994] IFLR (March) 31 at 32; LC Buchheit, *How to Negotiate Eurocurrency Loan Agreements* 2nd edn, (London, Euromoney Publications, 2000), p 110.

made or continued within the terms of the agreement. There is a risk that substantial damages may be awarded against the bank.

If we assume that the bank is in breach of contract for failing to make the advance or calling in a loan prematurely, what remedies are available to the borrower? An award of specific performance against the bank is unlikely. This follows from *South African Territories Ltd v Wallington*,[98] where House of Lords refused to grant specific performance of a contract to lend money. The general rule is that specific performance will not be ordered where damages are an adequate remedy. However, in *Loan Investment Corporation of Australasia v Bonner*,[99] the Privy Council said that 'in exceptional circumstances' specific performance might be awarded in the case of an unsecured loan agreement. There was no elaboration, but this might apply where there were difficult questions about the measure and remoteness of damages or obviously great delay and expense in obtaining them. There is an argument that specific performance could be ordered if the loan agreement is regarded as a 'debenture' as to fall within s 195 of the Companies Act 1985. The section provides that a contract with a company to take up and pay for debentures of a company may be enforced by an order for specific performance. The term 'debenture' has no hard and fast meaning, although it has been generally described as encompassing any document which creates or acknowledges a debt.[100] However, a loan agreement probably does not do this as the debt is not actually created until after the agreement is concluded and the money advanced.[101]

This leaves the borrower with a claim in damages for breach of contract. Damages for breach of contract are designed to compensate for the damage, loss or injury the claimant has suffered through the breach. The object of an award of damages for breach of contract is to place the claimant, so far as money can do it, in the same position, with respect to damages, as if the contract had been performed.[102]

Three issues must be kept in mind. The first is causation. In order to establish a right to damages the claimant must show that the breach of contract was the 'effective' cause of the loss. The courts usually deal with this issue by turning to their 'common sense' in interpreting the facts.[103] Secondly, there is remoteness. Certain losses may be too 'remote' and, therefore, irrecoverable. The classic test of remoteness is to be found in the judgment of Alderson B in *Hadley v Baxendale*.[104] The test can be divided into

[98] [1898] AC 309.
[99] [1970] NZLR 724.
[100] *Levy v Abercorris Slate and Slab Co* (1887) 37 Ch D 260 at 264, Chitty J.
[101] E Ferran, *Company Law and Corporate Finance* (Oxford, OUP, 1999), pp 463–4; RC Tennekoon, *The Law and Regulation of International Finance* (London, Butterworths, 1991), pp 125–7; cf A Berg, 'Syndicated Lending and the FSA' [1991] IFLR (January) 27.
[102] *Robinson v Harman* (1848) 1 Exch 850 at 855.
[103] *Galoo Ltd v Bright Grahame Murray* [1994] 1 WLR 1360, 1374–5.
[104] (1854) 9 Exch 341 at 354.

two limbs:[105] (*i*) the defendant is liable for such losses as occur 'naturally' or as a result of the 'usual course of things' after such a breach; and (*ii*) the defendant is liable for losses which did not arise 'naturally' but were within the reasonable contemplation of both the parties at the time they made the contract. In both cases there must have been a 'serious possibility' or a 'real danger' or a 'very substantial' probability that the loss would occur as a result of the breach.[106] Exceptional losses, such as the highly lucrative Ministry of Supply contracts that were lost as a result of the defendant's failure to deliver the boiler in *Victoria Laundry (Windsor) Ltd v Newman Industries Ltd*,[107] can only be recovered under the second limb of the rule.[108] Liability under the second limb depends on the defendant's knowledge, acquired before or at the time he entered into the contract, of those special circumstances that mean that a breach is liable to cause more loss than would arise in the ordinary course of things. That the defendant must at least have knowledge of the special circumstances is clear,[109] but there is some authority for the view that, in addition, the claimant must show that the defendant agreed to assume liability for the exceptional loss.[110] Finally, there is the question of mitigation. The claimant must take reasonable steps to mitigate his loss. The claimant cannot claim compensation for loss which is not due to the breach but to his own failure to behave reasonably after the breach. However, damages will not be reduced where failure to mitigate is due to the claimant's impecuniosity.[111]

It is submitted that when these general principles are applied to the bank's breach of contract in failing to advance funds to the borrower and/or terminating the facility and calling in the loan, the following proposition can be made:

(a) The borrower will recover little more than nominal damages where he can secure broadly the same lending terms from another lender (assuming he has suffered no other loss).[112] Nevertheless, the cost of negotiating the substitute loan may be recoverable.[113] This seems to fall within the first limb of the rule in *Hadley v Baxendale*.

[105] Although there are two limbs to the rule in *Hadley v Baxendale*, they do in fact form part of a single general principle.

[106] *Koufos v C Czarnikow Ltd* (*The Heron II)* [1969] 1 AC 350, 388, 414–5, 425.

[107] [1949] 2 KB 528.

[108] And in the *Victoria Laundry* case they were held not to be recoverable because the defendant had not been given any information about these particular contracts.

[109] *Patrick v Russo-British Grain Export Co Ltd* [1972] 2 KB 535 at 540.

[110] *Horne v Midland Railway* (1873) LR 6 CP 131 at 141. For criticism of this view, see J Beatson, *Anson's Law of Contract* 28th edn, (Oxford, OUP, 2002), p 609.

[111] *Clippens Oil Co Ltd v Edinburgh & District Water Trustees* [1907] AC 291, 303; *Alcoa Minerals of Jamaica Inc v Herbert Broderick* [2000] 3 WLR 23, 30–31, PC. See also B Coote, 'Damages, *The Liesbosch*, and Impecuniosity' [2001] CLJ 511.

[112] *Manchester & Oldham Bank v Cook* (1883) 49 LT 674 at 678, Day J.

[113] *Prehn v Royal Bank of Liverpool* (1870) LR 5 Exch 92.

(b) But where the cost of the alternative finance is greater than the original loan the borrower will recover the difference. In *South African Territories Ltd v Wallington*, Chitty LJ stated:[114]

> The plaintiffs are entitled to recover damages for breach of the agreement to make the loan. The damages in such case may be large or small, or merely nominal, according to the circumstances. The measure of the damage is the loss sustained by the borrower through the breach … If the intended borrower, being a man of good credit, can readily obtain the loan from another person on the same terms, the damages would be nominal. If he cannot obtain the money except at a higher rate of interest, or for a shorter term of years, or upon other more onerous terms, the damages would be greater and might be very substantial. The burden of proof of proving the amount of the loss suffered rests on the plaintiff.

The cost of the alternative finance may be greater than the original finance because of (a) a deterioration in the borrower's own position, and/or (b) general market deterioration. It has been said that a lender is not responsible for extra costs of a new loan 'which are attributable to a decline in the borrower's credit in the interim'.[115] In *Bahamas (Inagua) Sisal Plantation v Griffin*,[116] the claimant company sought to recover damages from the defendant for his breach of contract to subscribe for debentures in the company. The claim for damages failed as it was held that the cause of the company's failure to find another person to take the debentures was that 'the company had fallen into disrepute and bad financial odour, the defendant was not responsible for that'.[117] However, it is submitted that where the deterioration in the borrower's position is due to the fact that the bank has wrongfully withdrawn or failed to advance funds then the extra cost of the alternative finance should be recoverable as within the contemplation of the parties at the time the contract was made. Similarly, where general deterioration in the market increases the cost of borrowing, such loss should be recoverable as within the reasonable contemplation of the parties, unless the market deterioration is so catastrophic that is goes beyond that which could be reasonably contemplated, which would be rare. *South Australia Asset Management Co v York Montague Ltd*, where the House of Lords held that valuers were not liable for loss

[114] [1897] 1 QB 692 at 696–7, CA; affd [1898] AC 309.
[115] PR Wood, *International Loans, Bonds and Securities Regulation* (London, Sweet & Maxwell, 1999), para 2–8, citing *Bahamas (Inagua) Sisal Plantation v Griffin* (1897) 14 TLR 139.
[116] (1897) 14 TLR 139, Bigham J.
[117] *Ibid*, at 140.

attributable to the collapse of the property market, turned on the scope of the duty undertaken and so is distinguishable in this respect.[118]

(c) Terminating the bank's lending commitment and acceleration of repayment of the loan could trigger cross-default provisions in other financing and result in the bankruptcy of the borrower. Potential liability for the bank could be huge and extend to the loss of the borrower's going concern value. Again, such loss should be within the first limb of the rule in *Hadley v Baxendale* as cross-default clauses are common in finance documents and should be within the reasonable contemplation of the parties. *Crimpfil Ltd v Barclays Bank plc* illustrates that the borrower's claim could be sizeable.[119] The claimant, a company trading in fabric and textiles, brought an action against the defendant bank for breach of contract as a result of the termination of substantial facilities offered to the claimant. The disruption to the claimant's trade was so severe that it was forced into administration and eventual closure. Liability was admitted for wrongful termination of the main facility, and judgment given for the claimant in respect of the remaining facilities. The quantum of damages was in dispute. The claimant contended that the business, which was undergoing change and therefore vulnerable to action by the bank, was basically healthy, and so claimed damages of £6 million. The bank maintained that the business was bound to fail and consequently the bank's action merely hastened the claimant's demise so that the loss was nil. On the claimant's application for interim damages the judge awarded £1.8 million on the basis that the damages likely to be recovered at the hearing on quantum was in the region of £3 million. The Court of Appeal upheld the judge's award.

(d) Suspending loans temporarily could also result in the failure of the business if such funds were necessary to maintain operations. On the other hand, the borrower must take reasonable steps to mitigate his loss, which might include using cash in hand, reducing expenses, deferring construction activities, curtailing capital expenditure and seeking other sources of funds. In this case, damages might be awarded for foregone business opportunities or the cost of alternative funding.[120] These losses could be less that those arising from collapse of the borrower's business but they may still be substantial.

[118] [1997] AC 191.
[119] [1995] CLC 385.
[120] RM Gray, 'Does the crisis bring default under MAC clauses?' [1998] IFLR (April) 17 at 19.

(e) Suspending loans temporarily and termination of the facility can lead to the borrower losing profit on particular business opportunities which he can no longer exploit. Whether such loss of profit is recoverable depends on whether it was within the reasonable contemplation of the parties at the time the contract was made. The borrower would have to bring himself within the second limb of *Hadley v Baxendale* and show that the bank had sufficient notice of these opportunities and (possibly) that it agreed to assume liability for the exceptional loss. Where the loan has been made for a specific purpose—acquisition finance or project finance—the bank could well find itself liable for such loss. In *Manchester and Oldham Bank Ltd v WA Cook*,[121] the bank agreed to make a loan which the borrower required to enable him to purchase an interest in a colliery. The bank knew the purpose of the loan. The deal collapsed when the bank defaulted.[122] The loss of profit from the colliery purchase was held recoverable.

In summary, there is a real risk that, in some circumstances, the bank's incorrect invocation of a MAC clause to suspend or terminate the loan facility may result in a large award of damages being made against it.

XI. CONCLUSION

The MAC clause is essentially a self-help provision inserted into a financial agreement to deal with the unpredictable. Its form is influenced by one pragmatic consideration: the need to get the deal done. But there is a price to be paid. The MAC clause is generally wide in scope and vague. This leads to uncertainty as to when the clause can be relied upon. Because of this the general view is that it would be foolish to rely exclusively on the MAC clause as the basis for accelerating repayment of a loan.[123] The events of 9/11 may have caused banks and their lawyers to look at the issue again, but it is unlikely that that advice will change. The potential liability in damages for getting it wrong is simply too great to warrant the risk.

Nevertheless, the MAC clause is not worthless surplusage. It may have suffered a blow in the context of takeovers of UK public companies following WPP/Tempus, but it still fulfils a useful commercial function in other areas.

[121] (1883) 49 LT 674. See also *Astor Properties v Tunbridge Wells Equitable Friendly Society* [1936] 1 All ER 531.
[122] It seems that the borrower was unable to obtain the money elsewhere.
[123] See, eg, R Youard, 'Default in International Loan Agreements I and II' [1986] JBL 276 and 378 at 390; RC Tennekoon, *The Law and Regulation of International Finance* (London, Butterworths, 1991), p 99; LC Buchheit, *How to Negotiate Eurocurrency Loan Agreements* 2nd ed, (London, Euromoney Publications, 2000), p 110.

At the very least the MAC clause usually provides a party with some bargaining power should things turn sour.[124] Banks, for example, can more easily refuse to advance funds previously committed to a borrower who then gets into difficulties if absence of material change is a condition to further lending.[125] There are risks of course, but these may be outweighed by the fact that the deal (or, in this case, the restructured deal) gets done.

[124] David Morley, 'Clauses with potent effects', *Financial Times*, 4 March 2002.
[125] *Ibid*.

12

Re-thinking Insurable Interest

JOHN LOWRY AND PHILIP RAWLINGS*

I. INTRODUCTION

THE PURPOSE OF property insurance is to shift the risk of loss from the insured to the insurer. It is, therefore, axiomatic that the insurer should only be liable to indemnify the insured for the loss suffered.[1] Superimposed on this principle of indemnity is the requirement of insurable interest. This produces the curious result that two apparently separate and distinct fundamental principles of insurance law are harnessed in order to serve one objective, namely, to determine the liability of the insurers for the loss suffered. In the mid-eighteenth century there seems to have been some uncertainty as to whether an insurable interest was, indeed, required for an enforceable policy. Certainly the courts were enforcing not just contracts in which the insurer agreed to be liable 'interest or no interest', but, more generally, wagering agreements.[2] However, anxieties about the perceived evils inherent in wagering together with the concern that allowing those who lack interest to insure might encourage them to bring about the loss led to legislative intervention. This process began with the Marine Insurance Act 1745,[3] which laid down as a prerequisite to the validity of certain insurance contracts that the

* We thank Professor Malcolm Clarke of the University of Cambridge and Dr James Penner of the LSE for their comments on an earlier draft. We also owe a debt of gratitude to Sir Jonathan Mance and Adrian Hamilton QC for their incisive commentaries on this chapter at the LSE Seminar.

[1] A policy is not prevented from being an indemnity merely because it only covers part of the loss. Valued policies are not contracts of indemnity in the strict sense since the parties agree the measure of loss at the time that the contract is concluded rather than at the time of the loss.

[2] Although wagers were held not to be enforceable in *Goddert v Garrett* (1692) 2 Vern 269, they were enforced in *Assevieda v Cambridge* (1710) 10 Mod 77; *Harman v Van Hatton* (1716) 2 Vern 717; *De Paiba v Ludlow* (1721) 1 Comyns 361; *Dean v Decker* (1746) 2 Str 1250. See also *Craufurd v Hunter* (1798) 8 TR 13 at 23, per Lord Kenyon CJ, at 25, per Grose J; W Beavis, *Lex Mercatoria Rediviva: or, The Merchant's Directory*, (London: Edward Comyns, 1752), at 266; D MacLachlan (ed), *Arnould on the Law of Marine Insurance*, (London: Stevens & Sons, 1872), v I, at 112–13; JA Park, *A System of the Law of Marine Insurances*, (London: A Strahan, 1800), at 259. See generally, R Merkin, 'Gambling by Insurance—A Study of the Life Assurance Act 1774' [1981] *Anglo-American LR* 331.

[3] 19 Geo II, c 37.

insured possess an insurable interest in the subject matter of the policy.[4] This left the question open as to how insurable interest was to be defined.

This chapter considers the competing tests for determining the requirement of insurable interest: on the one hand, a broad conception, which has come to be known as the factual expectation test, whereby the determinative question is whether or not the insured suffered some loss from damage to the subject matter, or stood to gain some advantage from its continued existence; and, on the other hand, a narrower conception, referred to as the legal interest test, whereby the insured is required to demonstrate some legally enforceable right in the insured property. After some uncertainty, the latter emerged as the orthodox approach. Yet, it is our contention that the policy concerns which serve to underpin the legal interest test no longer prevail. Indeed, it will be argued that insistence on this test can operate as a trap for the unwary in so far as it renders a policy void thereby frustrating reasonable commercial expectations. Further, it is questionable whether modern insurance law should facilitate the evasion of obligations freely entered into by insurers with full knowledge. These problems have long been recognised in many other common law jurisdictions; as a result, they have adopted the factual expectation test. Moreover, in a series of modern decisions the English courts appear to be recognising the need for a broader conception of insurable interest. However, we will go on to argue that if it is accepted both that the original policy reasons for insurable interest no longer apply and that the underlying purpose of insurance is to shift the risk of pecuniary loss, then the requirement for insurable interest can be dispensed with altogether because its functions are effectively discharged by the principle of indemnity.

In this chapter, we consider the emergence of the principle of insurable interest culminating in the decision of the House of Lords in *Macaura v Northern Assurance Co*.[5] We then go on to examine the rather different approach taken by judges in other common law jurisdictions. Finally, we discuss how modern English courts have reassessed insurable interest in a series of subrogation cases.

II. THE EMERGENCE OF THE INSURABLE INTEREST REQUIREMENT

Like many other eighteenth-century statutes, the preamble to the Marine Insurance Act 1745, in which its purpose is set out, is somewhat obscure:

> the making of assurances, interest or no interest, or without further proof of interest than the policy, hath been productive of many pernicious practices,

[4] See s 1.
[5] [1925] AC 619.

> whereby great numbers of ships, with their cargoes, have either been fraudulently lost and destroyed, or taken by the enemy in time of war; and such assurances have encouraged the exportation of wool, and the carrying on many other prohibited and clandestine trades, which by means of such assurances have been concealed, and the parties concerned secured from loss, as well to the diminution of the publick revenue, as to the great detriment of fair traders: and by introducing a mischievous kind of gaming of wagering, under the pretence of assuring the risque on shipping, and fair trade, the institution and laudable design of making assurances, hath been perverted; and that which was intended for the encouragement of trade and navigation, has in many instances, become hurtful of, and destructive to the same.[6]

By section 1 all insurance contracts on British ships and their cargoes are declared 'null and void' where made 'interest or no interest, free of average, and without benefit of salvage to the assurer',[7] even if the 'insured' does, in fact, have an insurable interest. In a series of early cases in which the scope of the statute was considered, Lord Mansfield CJ expressed the view that it made 'insurance a contract of indemnity'.[8] For instance, in *Lowry v Bourdieu,*[9] the claimants, who had lent money to a ship's captain on the security of his bond, took out an insurance policy on his ship and its cargo. Lord Mansfield held the policy to be a wager and, therefore, void: 'it was a hedge. But they had no interest; for, if the ship had been lost and the underwriters had paid, still the plaintiffs would have been entitled to recover the amount of the bond.'[10] He went on to observe that:

> There are two sorts of policies of insurance; mercantile and gaming policies. The first sort are contracts of indemnity, and of indemnity only … . The second sort may be the same in form, but in them there is no contract of indemnity, because there is no interest upon which a loss can accrue. They are mere games of hazard; like the cast of a die.[11]

Significantly, Lord Mansfield was not denying the enforceability of wagering contracts generally, but merely acknowledging that the 1745 Act made

[6] See also, 43 Eliz, c 12, preamble; W Beavis, above, n 2; N Magens, *An Essay on Insurances*, (London: W Baker, 1755), at 2; S Marshall, *A Treatise on the Law of Insurance*, (London: A Strahan, 1802), v I, at 3; R Lowndes, *Insurable Interest and Valuations*, (London: Stevens & Sons, 1884).

[7] See *Kent v Bird* (1777) 2 Cowp 583. The provisions in the 1745 Act were repealed by the Marine Insurance Act 1906 (s 92) and replaced by s 4 under which a wager policy is void, but not illegal (but see Marine Insurance (Gambling Policies) Act 1909).

[8] *Le Cras v Hughes* (1782) 3 Dougl 79, at 86. See also, *Moran, Galloway & Co v Uzielli* [1905] 2 KB 555 at 563, per Walton J.

[9] (1780) 2 Dougl 468.

[10] *Ibid,* at 470. See also *Moran, Galloway & Co v Uzielli*, above, n 8.

[11] Above n 9, at 470. It is worth noting here that this did not eradicate the practice of parties entering into what are known as p p i (policy proof of interest) contracts: the insurers do not raise the issue of the insurable interest as a defence to a claim, although if the policy is the subject of litigation over another issue the court will refuse to enforce it. See further, *Gedge v Royal Exchange Assurance Co* [1900] 2 QB 214.

them unenforceable in relation to marine adventures by introducing the requirement that the insured demonstrate an insurable interest.

A general definition of insurable interest proved elusive. In *Le Cras v Hughes*,[12] Lord Mansfield, rather unhelpfully, stated that: '[a]n interest is necessary, but no particular kind of interest is required.'[13] He did go on to stress, however, that it was not necessary to possess a legal interest in the insured property. Briefly, the issue was whether the crews of a Royal Navy squadron had an insurable interest in two enemy ships they had seized. Lord Mansfield decided that they had for two reasons. The first was that the crews had rights vested in them by the Prize Act which gave them a sufficient interest to support the insurance. This was uncontroversial, but he went on to justify the decision on the separate ground that: '[w]herever a capture has been made, since the Revolution, by sea or land, the Crown has made a grant [of the prize ships]: there is no instance to the contrary.'[14] Although the sailors had no legal right to the prize, he nevertheless regarded the moral certainty of the Crown granting it to them as sufficient for an insurable interest. Similarly, in *Boehm v Bell*,[15] it was held that those who capture a ship as a prize may take out insurance because their possession brought with it certain rights (if it were judged a legal capture, the crew would be entitled to the prize) and duties (if it were an improper capture, the crew might be liable for damages) and they, therefore, had an insurable interest. Unfortunately, in that case Grose J's discussion of the general requirement of insurable interest added little beyond stating that:

> the whole difficulty has arisen from confounding an absolute indefeasible interest with an insurable interest. It is not pretended that the assured had the absolute property in the subject of insurance; neither need they have such property to make the policy legal; it is sufficient if they had an insurable interest.[16]

The late eighteenth-century case law on insurance contracts not covered by the 1745 Act seems principally to have been driven by the notion that agreements should be enforced rather than be defeated by the anxiety over wagering.[17] The view was taken that insurance should be encouraged because it played a key role in commerce by enabling people with little capital to engage in business by reducing their exposure to risk and ruin:[18] as Marshall observed in his textbook of 1802, 'insurances are made for the

[12] Above n 8.
[13] *Ibid*, at 86.
[14] *Ibid*.
[15] (1799) 8 TR 154.
[16] *Boehm v Bell* (1799) 8 TR 154 at 161.
[17] This was in spite of the efforts of judges such as Buller J: see, *Atherfold v Beard* (1788) 2 TR 610 and *Good v Elliott* (1790) 3 TR 693. See further, R Merkin, above n 2, at 334–5.
[18] See generally, W Beavis, above n 2; R Lowndes, above n 6; N Magens, above n 6; S Marshall, above n 6, v I, at 99–100.

encouragement of trade'.[19] This emphasised the importance of focusing on the social and commercial benefits of insurance rather than on the limitations imposed by legislation on wagering. Nevertheless, it was the policy against wagering underlying the 1745 Act that proved to be the decisive issue in the leading decision of *Lucena v Craufurd*.[20]

In *Lucena* the issues closely mirrored those that had confronted the courts in *Le Cras* and *Boehm*. Although the cases involved marine insurance, the views expressed have been applied more generally to the definition of insurable interest. Briefly, the facts were that commissioners had been given statutory authority to take charge of Dutch ships and cargoes in England. Acting under the orders of the Admiralty, a Royal Navy ship took several Dutch ships at sea. The commissioners then arranged to insure the ships while they were on their way to England. They were lost before arrival. While the provisions of the legislation meant that the commissioners would clearly have had an insurable interest in the ships if they reached these shores, the issue was whether such an interest existed before their arrival. The case was first argued before Lord Kenyon CJ and a special jury at the Guildhall in 1799; it was then appealed through the Court of King's Bench and the Exchequer Chamber, and finally reached the House of Lords in 1802.

The overwhelming majority of the judges in the lower courts referred to the second ground given by Lord Mansfield for his decision in *Le Cras* as support for their view that the commissioners possessed an insurable interest.[21] For instance, in the Exchequer Chamber it was said that 'an inchoate interest though imperfect till a given contingency shall have taken place is nevertheless insurable'.[22] They went on to explain:

> It may be admitted that a mere hope or expectation cannot be insured, and it may therefore also be admitted that the next of kin to a person in a dying state and incapable of making a will, who has property on the sea, cannot insure that property .... Where nothing intervenes between the subject insured and the possession of it, but the perils insured against, the person so situated may insure the arrival of such subject of insurance, for he has an interest to avert the perils insured against.[23]

The judges' view was that once the Navy had taken the ships, the Crown acquired an insurable interest and, therefore, so did the commissioners who acted as trustees for the Crown: 'These commissioners had not a mere ideal

[19] S Marshall, above n 6, v I, pp 99–100. See also, R Lowndes, above n 6.
[20] (1806) 2 Bos & PNR 269; affirming (1802) 3 Bos B&P 75, considered *infra*.
[21] *Lucena v Craufurd* (1802) 3 B & P 75 at 90–9 *passim*; (1802) 2 B & P (NR) 269 at 294.
[22] (1802) 3 B & P 75 at 98.
[23] *Ibid*, at 94–5. The judges who subscribed to this view in the Exchequer Chamber were Graham B, Rooke J, Thompson B, Hotham B, Macdonald CB, Lord Avanley CJ and Heath J.

expectation of probable interest, but an interest vested in them as trustees.'[24] The judges denied that there was any significance in the fact that the legal interest in the ships only vested in the commissioners upon their arrival in England. They reasoned that others were in a similar position but still had an insurable interest, such as the ship owner who could insure the freight even though the right to freight was contingent on the ship reaching its destination.

Similarly, the majority of the judges called to give their advice to the House of Lords argued that: 'a vested interest is not necessary to give the right of insuring. The commissioners had a contingent interest; and supposing the intentions of the Crown to remain unaltered, nothing stood between them and the vesting of that contingent interest but the perils insured against'.[25] The judges summed up their approach:

> The question always is, whether the policy be a gaming contract? If it be no artifice how can it elude the force of the statute? The case of *Le Cras v Hughes* was infinitely more likely to introduce an abuse of the statute than the present case. That has been decided above 20 years; yet what ill consequences have followed? The same may be said of valued policies. In the case of wagering policies, any number of persons may make insurances on the same ship. But that is not the case here. If the commissioners could not insure this property, the Dutch owners could not; and it would be a strange paradox to assert, that these are ships and cargoes subject to all the perils of the sea in their voyage, and yet none are competent to insure them.[26]

Before the House of Lords, two judges dissented from the advice offered by their colleagues and asserted that there was no insurable interest. Chambre J, who had also dissented in the Exchequer Chamber, stated that the statute appointing the commissioners did not give them powers with regard to property until it arrived in the country, so that the only claim of interest was 'a mere naked expectation of acquiring a trust or charge respecting the property without a scintilla of present right either absolute or contingent, in possession, reversion, or expectancy, in the proper legal sense of the word.'[27] While Lawrence J similarly denied the existence of insurable interest, his reasoning differed. It is worth considering his opinion at greater length because of its enduring influence. He began by defining the nature of an insurance contract in terms of the protection it afforded insured parties

[24] *Ibid*, at 98.
[25] Before the House of Lords, Le Blanc J, Grose J and Sir James Mansfield CJ took the same view as the majority in the Exchequer Chamber (they were joined by judges that had already been of the majority in the lower court—Graham B, Rooke J and Heath J—and, with a reservation that did not undermine the general thread of the majority's argument, Thompson B): *Lucena v Craufurd* (1802) 2 B & P (NR) 269 at 289 *et seq*.
[26] *Ibid*, at 297.
[27] *Ibid*, at 299. Also, *Lucena v Craufurd* (1802) 3 B & P 75 at 100–5.

not merely against loss resulting in deprivation of property but also against uncertain events which may lead to some other disadvantage, such as loss of anticipated profit. The risk of such loss, damage or other prejudice is thereby shifted to the insurer.

> That a man must somehow or other be interested in the preservation of the subject-matter exposed to perils, follows from the nature of this contract, when not used as a mode of wager, but as applicable to the purposes for which it was originally introduced; but to confine it to the protection of the interest which arises out of property, is adding a restriction to the contract which does not arise out of its nature … . Interest does not necessarily imply a right to the whole, or a part of a thing, nor necessarily and exclusively that which may be the subject of privation, but the having some relation to, or concern in the subject of the insurance, which relation of concern by the happening of the perils insured against may be so affected as to produce a damage, detriment, or prejudice to the person insuring: and where a man is so circumstanced with respect to matters exposed to certain risks or dangers, as to have a moral certainty of advantage or benefit, but for those risks or dangers he may be said to be interested in the safety of the thing.[28]

Having spoken generally about the nature of interest and insurance, Lawrence J went on to formulate what has become known as the factual expectation test:

> To be interested in the preservation of a thing, is to be so circumstanced with respect to it as to have benefit from its existence, prejudice from its destruction. The property of a thing and the interest deviseable from it may be very different: of the first the price is generally the measure, but by interest in a thing every benefit and advantage arising out of or depending on such thing, may be considered as being as comprehended.[29]

This represented a more precise formulation of the views he had expressed in earlier cases. For instance, in 1799 he spoke of the need to ask if there was 'a loss against which they might indemnify themselves, by a policy of insurance? I do not mean a certain, but a possible loss.'[30] In *Lucena*, he seems to require more than a mere possibility, so that, in contrast to his colleagues, he declined to follow Lord Mansfield's approach in *Le Cras* and concluded that the commissioners had no insurable interest. In his view, they were not trustees because they had no 'existing right to the thing insured for the benefit of another',[31] nor were they consignees in whom a

[28] (1802) 2 B & P (NR) 269 at 301–2.
[29] *Ibid*, at 302–3. For an earlier and less developed expression of his views, see *Barclay v Cousins* (1802) 2 East 545.
[30] *Boehm v Bell*, above n 16, at 162. See also, *Barclay v Cousins*, above n 29.
[31] Above n 20, at 304.

right had been vested by a bill of lading or similar instrument. He took the view that the commissioners only had an expectation of an expectation:

> if it shall appear from [the commission] that the purpose and object ... was only to take care of the Dutch property after its arrival in England, and if till then they had not any power to interfere with it, they cannot be said at the time of the sailing insurance and loss to have been interested; until the time should arrive when their authority and duty as commissioners would attach, they would have no existing concern in such property, and could not in their character as commissioners suffer any prejudice or damage before they had any concern in the thing assured.[32]

In the House of Lords, Lord Eldon delivered the leading speech. Although he agreed with Lawrence J that there was no insurable interest, he was keen to emphasise the very different reasoning which led him to that conclusion. His main concern was with wagering. Curiously, he rejected the suggestion that before the 1745 Act insurance might have been effected without interest, but, in any event, he took the view that the Act was decisive and that the courts should follow its spirit: 'lest that sort of wagering in policies should grow up, which has of late been extending itself considerably.'[33] He rejected the alternative ground for the decision in *Le Cras*, namely, that expectation of a grant by the Crown was sufficient: '[t]hat expectation, though founded upon the highest probability, was not interest, and it was equally not interest, whatever might have been the chances in favour of the expectation.'[34] While some argued it was possible to insure where the interest amounted only to an 'intermediate thing between a strict right, or a right derived under a contract, and a mere expectation or hope',[35] Lord Eldon, 'in vain endeavoured ... to find a fit definition of that which is between a certainty and an expectation'.[36] He concluded by stressing that he was unable to pinpoint with any degree of precision 'what is an interest unless it be a right in the property, or a right derivable out of some contract about the property, which in either case may be lost upon some contingency affecting the possession or enjoyment of the party.'[37] Lord Eldon went on to hypothesise:

> Suppose A to be possessed of a ship limited to B in case A dies without issue; that A has 20 children, the eldest of whom is 20 years of age; and B is 90 years of age; it is a moral certainty that B will never come into possession,

[32] *Ibid*, at 305.
[33] *Ibid*, at 323.
[34] *Ibid*, at 323. See also *Routh v Thompson* (1809) 11 East 428 at 433, per Lord Ellenborough CJ.
[35] Above n 20, at 321.
[36] *Ibid*.
[37] *Ibid*, at 321.

> yet there is a clear interest. On the other hand, suppose the case of the heir at law of a man who has an estate worth £20,000 a year, who is 90 years of age; upon his death-bed intestate, and incapable from incurable lunacy of making a will, there is no man who will deny that such an heir at law has a moral certainty of succeeding to the estate; yet the law will not allow that he has any interest, or anything more than a mere expectation.[38]

He noted that in the prevailing legislation governing insurance: 'the objects of insurance are plainly described to be ships, cargoes, wares, mechandizes, or effects.'[39] There was nothing showing that 'expectation of profits and some other specifies of interest which have been insured in later times' were insurable. He did not overrule post-1745 Act decisions that suggested these types of interest were insurable, but he did undermine their authority and certainly was keen not to see them extended. So, for instance, he disagreed with the suggestion that an expectation of unascertained profits from a marine adventure might be insured. It was his view that the virtue of a restrictive approach lies in what he saw as its inherent certainty: '[C]ourts of justice sit here to decide upon rights and interests in property … They do not sit here to decide upon things in speculation.'[40] For Lord Eldon the appropriate test was to ask whether the insured possessed 'a right in property, or a right derivable out of some contract about the property, which in either case may be lost upon some contingency affecting the possession or enjoyment of the party?'[41] On this basis he said of the commissioners:

> If they have a right so to insure, it seems to me that any person who is directed to take goods into his warehouse may insure; and that there is nothing to prevent the West India Dock Company from insuring all the ships and goods which come to their docks. If moral certainty be a ground of insurable interest, there are hundreds, perhaps thousands, who would be entitled to insure. First the dock company, then the dock-master, then the warehouse-keeper, then the porter, then every other person who to a moral certainty would have any thing to do with the property, and of course get something by it.[42]

Whichever of the two tests—Lawrence J's factual expectation test, or Lord Eldon's legal interest test—is applied, the final outcome in the majority of cases will be the same. The decision in *Lucena* itself illustrates the point. In his discussion of whether the Dutch commissioners were similar to consignees in whom an interest is vested by some instrument of consignment,

38 *Ibid*, at 325.
39 *Ibid*, at 321.
40 *Ibid*, at 325.
41 *Ibid*, at 321.
42 *Ibid*. Following *Lucena*, it was clear that in circumstances, such as those in *Le Cras*, where the ship is taken as a prize, then legal interest vested in the crown at the time of capture, but not in the captors, until conferred by the crown: contrast *Routh v Thompson* (1809) 11 East 428 with *Routh v Thompson* (1811) 13 East 274.

such as a bill of lading, Lawrence J said that if they were consignees they were 'naked consignees who have not the legal property of the subject-matter of the insurance, and who are not beneficially interested in it.'[43] As such they had no insurable interest because 'such consignee can himself suffer no prejudice by the total or partial destruction of a thing which forms no part of his property.'[44] Put simply, not only did they lack any strict proprietary interest in the ships, they were also unable to establish a sufficient economic interest to support a claim of 'moral certainty of advantage or benefit' from their continued existence.

Notwithstanding the depth of Lord Eldon's reasoning, in subsequent cases heard during the nineteenth century the judges, in so far as the definition of insurable interest was considered, expressed support for factual expectancy. For example, in *Lloyd v Fleming*,[45] Blackburn J said:

> This subject-matter [of the insurance] need not be strictly a property, in either the ship, goods, or freight; for, as has been long said, if a man is so situated with respect to them that he will receive benefit from their arriving safely at the end of the adventure, or sustain loss in consequence of their not arriving safely, he has an insurable interest; see *per* Lawrence J, in *Lucena v Craufurd*.[46]

The state of confusion over the proper test for determining insurable interest can be seen in Walton J's judgment in *Moran, Galloway & Co v Uzielli*.[47] The judge began by endorsing the broad conception of interest before switching midway to the narrower legal interest test. He said that in *Lucena* Lawrence J had concluded that 'an interest which is not in the nature of a property legal or equitable in the things exposed to maritime perils may still be insurable'.[48] He thought it was 'unnecessary to cite cases to shew that this opinion of Lawrence J has always been accepted as a correct statement of the law.'[49] Walton J stressed that:

> Without citing later cases, it is sufficient to say that, in my opinion, the result of the authorities is that … an interest to be insurable is not necessarily a right, legal or equitable, in or charge upon or arising out of the ownership of the thing exposed to the risks insured against, and any interest may be insured which is dependent on the safety of the thing exposed to such risks … .[50]

[43] Above n 20, at 307.
[44] Above n 20, at 307.
[45] (1872) 7 LR QB 299. See also, Blackburn J's judgment in *Wilson v Jones* (1867) 2 LR Ex 139, at 150.
[46] (1872) 7 LR QB 299, at 302.
[47] Above n 8.
[48] *Ibid*, at 560.
[49] *Ibid*, at 561.
[50] *Ibid*, at 562.

He went on to refer with approval a passage from Chalmers' book on marine insurance to the effect that: 'The definition of insurable interest has been continuously expanding, and dicta in some of the older cases, which would tend to narrow it, must be accepted with caution.'[51] Returning again to Lawrence J's approach, the judge concluded that to find that there was no insurable interest 'would, I think, be to impose an unnecessary fetter upon business which seems to me very ordinary and reasonable business, in no way tainted by the vice of wagering or gaming.'[52] Yet, his difficulty was that Lawrence J spoke in terms of 'interest dependent upon things exposed to the dangers to which maritime adventures are subjected', and this, he believed, required clarification. Walton J, therefore, concluded, in terms that resonate with Lord Eldon's approach:

> although an interest to be insurable is not necessarily a right, legal or equitable, in or charge upon or arising out of the ownership of the thing exposed to the risks insured against, and any interest may be insured which is dependent on the safety of the thing exposed to such risks, still it must in all cases at the time of the loss be an interest legal or equitable, and not merely an expectation, however probable.[53]

Notwithstanding the state of the case law, for marine insurance the issue appeared to be settled by Chalmers, who in drafting the Marine Insurance Act 1906, adopted Lord Eldon's test:

> a person is interested in a marine adventure where he stands in any legal or equitable relation to the adventure or to any insurable property at risk therein, in consequence of which he may benefit by the safety or due arrival of insurable property, or may be prejudiced by its loss, or damage thereto, or by the detention thereof, or may incur liability in respect thereof.[54]

In the non-marine context, the issue was settled by the House of Lords in *Macaura v Northern Assurance Co Ltd*.[55] Briefly, Macaura was the only substantial shareholder in a company to which he had sold timber on credit. He insured the timber in his own name and when it was destroyed by fire sought to claim against the policies. An initial allegation that Macaura's claim was fraudulent and dishonest was dismissed by an arbitrator and, apart from a brief statement to that effect, this was not mentioned in the House of Lords. The issue before the court was whether Macaura had an

[51] Sir Mackenzie Chalmers and Sir Douglas Owen, *A Digest of the Law Relating to Marine Insurance* (London: W Clowes & Sons, 1901), at 9.
[52] Above n 8, at 564.
[53] Above n 8, at 562.
[54] S 5(2).
[55] [1925] AC 619.

insurable interest in the timber owned by the company. Counsel for Macaura, drawing on Lawrence J, argued:

> Legal ownership is not necessary for insurable interest. So to confine it would be adding a restriction to a contract of insurance which does not arise out of its nature. To be interested in the preservation of a thing is to be so circumstanced with respect to it as to have benefit from its existence, prejudice from its destruction. If there is a legal certainty of loss arising from the destruction of the property insured then there is an insurable interest.[56]

The contention was that Macaura's insurable interest derived from his being the only shareholder. It was also argued that a separate insurable interest arose from his being the only substantial creditor of the company whose only substantial asset from which the debts could be paid was the timber. On both grounds, it was claimed, Macaura was bound to benefit by the preservation of the timber and suffer by its destruction. All five of their Lordships disagreed,[57] ruling that Macaura had no insurable interest. In giving the leading speech, Lord Buckmaster cited, with approval, Walton J's reasoning in *Moran, Galloway & Co v Uzielli,*[58] in which the judge had said: 'in so far as the plaintiffs' claim depends upon the fact that they were ordinary unsecured creditors .... I am satisfied it must fail.'[59] Lord Buckmaster, therefore, dismissed the idea that a creditor had an insurable interest in the assets of a debtor. His principal objection to Macaura's contention, however, turned on his status and interest in the company as shareholder. If Macaura's argument was accepted, then in Lord Buckmaster's view each shareholder in every company would have an insurable interest in corporate assets and the extent of that interest 'could only be measured by determining the extent to which his share in the ultimate distribution would be diminished by the loss of the asset—a calculation almost impossible to make.'[60] Lord Buckmaster then explicitly attacked Lawrence J by saying, 'I find ... difficulty in understanding how a moral certainty can be so defined as to render it an essential part of a definite legal proposition.'[61]

Lord Sumner, proceeding on the basis that neither the company's debt to the insured nor his shares were exposed to fire, observed: 'the fact that he was virtually the company's only creditor, while the timber is its only asset, seems to me to make no difference ... he was directly prejudiced by the paucity of the company's assets, not by the fire.'[62] He stated that the insured 'stood in no 'legal or equitable relation to' the timber at all. He had

[56] *Ibid*, 622.
[57] Lords Buckmaster, Atkinson, Sumner, Wrenbury and Phillimore.
[58] Above n 8.
[59] *Ibid*, at 562.
[60] Above n 55, at 627.
[61] *Ibid*, at 627.
[62] *Ibid*, at 630.

no 'concern in' the subject insured. His relation was to the company, not to its goods ... .'[63] Further, as a gratuitous bailee he was not liable for the timber and, therefore, an interest could not be based on bailment.

In unequivocally rejecting the broad conception of insurable interest put forward by Lawrence J in favour of Lord Eldon's narrow test based upon proprietary interest, the House of Lords failed to grasp the opportunity of testing, from a modern standpoint, the continued viability of the policy considerations underlying the requirement. Other common law jurisdictions have not displayed such reticence.

III. FACTUAL EXPECTATION IN THE ASCENDANCY

In Canada, Australia and the USA a pragmatic approach has been adopted in response to the perceived social and commercial benefits which widespread insurance offers. The reasoning in *Macaura* has been rejected principally on the basis that an overly technical determination of the insurable interest requirement has the potential to defeat the reasonable commercial expectations of the parties.[64] In its place the courts have substituted factual expectancy as the determinative test. Thus, in *Constitution Insurance Company of Canada v Kosmopoulos*,[65] the facts of which closely resemble *Macaura*, the Supreme Court of Canada, which had previously followed the restrictive Eldon formulation,[66] overruled itself. Wilson J, for the majority, took the view that the definition of insurable interest was ripe for fundamental re-examination: 'if the application of a rule leads to harsh justice, the proper course to follow is to examine the rule itself rather than affirm it and attempt to ameliorate its ill effects on a case-by-case basis.'[67] She therefore refused to follow the expedient solution adopted by her colleague, McIntyre J, of distinguishing *Macaura* and piercing the corporate veil on the basis that this was a one-shareholder corporation.[68]

[63] *Ibid.*

[64] See further, A J Campbell, 'Some Aspects of Insurable Interest' (1949) 27 *Can Bar Rev* 1; D Galbraith, 'An unmeritorious defence—The requirement of insurable interest in the law of marine insurance and related matters' [1993] *Insurance LJ* 177; RA Hasson, 'Reform of the Law Relating to Insurable Interest in Property—Some Thoughts on *Chadwick v Gibraltar General Insurance*' (1983–84) 8 *Can Bus LJ* 114.

[65] (1987) 34 DLR (4th) 208. Noted by L Stuesser (1987–8) 13 *Can Bus LJ* 227.

[66] See, for example, *Guarantee Co of North America v Aqua-Land Exploration Ltd* (1965) 54 DLR (2d) 29. See also *Wandlyn Motels Ltd v Commerce General Insurance Co* (1970) 12 DLR (3d) 605.

[67] Above n 65, at 214.

[68] *Ibid*, 210. Although not cited by him, support for his approach can be found in *Durocher v Gévry* [1961] Que QB 283. The insured, the majority shareholder in Gévry Automobiles Inc, obtained insurance on a race horse which was registered as belonging to the company. Montgomery J pierced the corporate veil and found that the economic reality was that the insured owned the company and the company had not paid for the horse: 'it was a matter of little practical importance to the insurer whether the horse was registered in his personal name or in that of the company'.

Wilson J began by addressing the fundamental issue raised by the appeal, namely, the appropriate test for determining insurable interest. Reviewing Lord Eldon's reasoning, which had led him to reject the factual expectation test, she roundly dismissed as unpersuasive his opinion that a restrictive test for insurable interest was required. She was critical of the certainty perceived as one of the strengths of Lord Eldon's test when compared with that of Lawrence J, and in support of her view Wilson J cited a passage from Brown and Menzes, *Insurance Law in Canada*.[69] Commenting on *Macaura*, the authors conclude:

> After *Macaura,* it is no longer possible to claim merely that one would be adversely affected by the loss; the insured must assert that he owned an interest in the objects destroyed. This provides the illusion of great certainty. Property law is among the most technical and certain segments of the law. This certainty is totally illusory because the new formulation makes no concessions either to the reasons for which insurable interest is a component of insurance law or for commonplace business transactions .... Assuming that an insurable interest in 'things' must mean property, among the simple questions raised are matters such as how does one own a direct interest in property which is not in existence at the time of the contract? Can next season's crops or fluctuating inventory be insured? Are warehousing and other bailee policies subject to the law as set out in *Macaura* so as to limit the right to insure to the bailee's liability to the bailor?

With respect to Lord Eldon's anxiety that the adoption of the factual expectation test would lead to too much insurance, Wilson J concluded that that 'may also be illusory.'[70] Insureds are under a duty to disclose all material circumstances so that insurers can assess the risk and if an 'insurer cannot estimate the likelihood of the loss occurring (because, for example, the information is in the hands of third parties) then it does not have to write the policy.'[71] Recognising that the broader test may increase the liability of insurers upon the happening of a single event owing to an increase in policies for the same risk, she said:

> But insurance companies have always faced the difficult task of calculating their total potential liability arising upon the occurrence of an insured event in order to judge whether to make a particular policy or class of policies and to calculate the appropriate premium to be charged. It is not for this Court to substitute its judgment for the sound business judgment and actuarial expertise of insurance companies by holding that a certain class of policies should not be made because it will result in 'too much insurance.' I would have

[69] C Brown and J Menzes, *Insurance Law in Canada* (Scarborough, Ontario: Carswell, 1982), at 84.
[70] Above n 65, at 217.
[71] *Ibid*, at 218.

> thought that a stronger argument could be made that there is too little insurance … . A broadening of the concept of insurable interest would, it seems to me, allow for the creation of more socially beneficial insurance policies than is the case at present with no increase in risk to the insurer.[72]

Wilson J roundly rejected the argument that a broadly conceived notion of insurable interest would lead to an increase in the wilful destruction of insured property stating that the legal interest test provided no better deterrent against such moral hazard. She considered that insureds who have a legal or equitable interest would, in fact, have better access to the insured property and therefore more opportunity to destroy it than those with an interest in the broader sense: 'If Lawrence J's definition of insurable interest … were adopted, this moral hazard would not be increased. Indeed, the moral hazard may well be decreased because the subject-matter of the insurance is not usually in [their] possession or control.'[73] Recognising that there might be an incentive to sole shareholders to destroy corporate assets if insurance moneys were paid to them free of the company's creditors, she pointed to company law doctrines, including the constructive trust and directors' duties, by which the courts can make the proceeds of insurance held by a shareholder available to the company. By such means the shareholder is prevented from benefiting personally.

Addressing Lord Buckmaster's concern regarding the difficulty of valuing a shareholder's interest in a company, Wilson J did not see this as insurmountable. She pointed to provisions in the Ontario Business Corporations Act, which require, in certain circumstances, such valuations to be made. Having laid such technical objections to rest, Wilson J surveyed the jurisprudence from a number of American jurisdictions which have rejected the restrictive definition of insurable interest in favour of factual expectancy and concluded that: 'No material has been referred to us by counsel to show that these developments in the United States have led to insoluble problems of calculation, difficulties in ascertaining insurable interests, wagering, over-insurance or wilful destruction of property.'[74] It is, perhaps, worth noting at this point that the English courts routinely award share purchase orders to petitioners claiming unfair prejudice under section 459 of the Companies Act 1985.[75] In requiring majority shareholders to purchase the minority shareholders' shares, the English courts have been able to frame a number of rules governing the valuation process.[76]

[72] *Ibid*, at 218.
[73] *Ibid*, at 224.
[74] *Ibid*, at 227.
[75] See the Companies Act 1985, s 459 and 461(2)(d).
[76] See, for example, *Re Bird Precision Bellows Ltd* [1984] Ch 419, [1986] Ch 658, CA. See also, *O'Neill v Phillips* [1999] 1 BCC 607, at 613–15 (Lord Hoffmann).

Wilson J's analysis reflects a shift in emphasis from Lord Eldon's concern, which led him to define insurable interest narrowly, to a view that recognises the economic and social benefits of insurance and, therefore, a broader conception of insurable interest. In the modern commercial world property insurance is generally sought to secure indemnification, and, as Wilson J points out, it is more socially beneficial to encourage widespread insurance than to restrict it. There seems, therefore, no convincing reason in this context for interfering with freedom of contract and, in particular, for not requiring insurers to meet liabilities under contracts which they have freely entered into and for which they have received premiums.

In Australia Lord Eldon's approach has also been discarded. The Australian Law Reform Commission (ALRC), in its report, *Insurance Contracts*,[77] concluded that Lawrence J's formulation 'would allow more flexibility to insurers and to the insuring public, without in any way promoting gaming and wagering in the form of insurance or adding to the risk of destruction of the property insured.'[78] In its opinion, technical rules had prevented the insured in *Macaura* from recovering the loss actually suffered by him. The ALRC considered that the strict legal interest test gave rise to results that were socially undesirable. For example, a named beneficiary under the will of a living testator stands to lose much of his projected inheritance if the testator's property is destroyed by fire. Yet, if the beneficiary effects a fire policy on the property, the legal interest test will prevent recovery notwithstanding actual loss. The ALRC also thought that the restrictive test produced commercially undesirable results. By way of example, it cited *Truran Earthmovers Pty Ltd v Norwich Union Fire Insurance Society Ltd*,[79] in which a purchaser of a bulldozer was held to have no insurable interest in the vehicle even though he had lent the owner money which was to be deducted from the purchase price: 'Once again, technical rules prevented recovery of an actual loss.'[80] The ALRC therefore proposed legislative reform to provide that 'where an insured is economically disadvantaged by damage to or destruction of the insured property, the insurer should not be relieved of liability by reason only that the insured did not have a legal or equitable interest in the property.'[81] This was given statutory effect by the Insurance Contracts Act 1984, section 17.[82]

[77] Report No 20 (1982), ch 5. See also the ALRC Discussion Paper No 63, *Review of the Marine Insurance Act* 1909 (2000), ch 7.
[78] Report No 20, above n 77 at para 120.
[79] (1976) 17 SASR 1.
[80] Report No 20, above n 77 at para 119.
[81] In reaching its recommendation for the adoption of economic interest as the determining question, the ALRC noted that several American States had adopted this test. It cited, by way of example, the Insurance Law New York s 158, which defines insurable interest as including 'any lawful and substantial economic interest in the safety or preservation of property from loss, destruction or pecuniary damage.'
[82] See also, the New Zealand Insurance Law Reform Act 1985, s 7(1) which abolishes the requirement of insurable interest in the case of life insurance and all contracts of indemnity insurance.

Although early US case law followed Lord Eldon's narrow formulation,[83] most states have now adopted economic interest as the determinative test.[84] Some thirty years before the decision in *Macaura*, the New York courts recognised that shareholders did have an insurable interest in corporate assets. In *Riggs v Commercial Mutual Insurance Co*,[85] it was held that stockholders 'have equitable rights of a pecuniary nature, growing out of their situation as stockholders, which may be prejudiced by the destruction of the corporate property … . It is very plain that both these rights of stockholders—viz, the right to dividends and the right to share in the final distribution of corporate property—may be prejudiced by its destruction.'[86] More broadly, at about this time the US Supreme Court, in *Harrison v Fortlage,*[87] decided that 'any person has an insurable interest in property, by the existence of which he will gain an advantage, or by the destruction of which he will suffer a loss, whether he has or has not any title in, or lien upon, or possession of the property itself.'[88] Similarly, in *National Filtering Oil Co v Citizens' Insurance Co* the court stated:[89]

> [A]n interest, legal or equitable, in the property burned is not necessary to support an insurance upon it; that it is enough if the insured is so situated as to be liable to loss if it is destroyed by the peril insured against; that such an interest in property connected with its safety and situation as will cause the insured to sustain a direct loss from its destruction, is an insurable interest;

[83] See, for example, *Farmers' Mutual Insurance Co v New Holland Turnpike Road Co* 122 Pa 37 (1888), where it was held that a turnpike company, which insured a bridge linking two sections of the company's road, lacked insurable interest because it had no legal or equitable interest in the bridge.
[84] See, for example, the decision of the Supreme Court of Massachusetts in *Hayes v Milford Mutual Fire Ins Co* 49 NE 754 (1898).
[85] 125 NY 7 (1890).
[86] *Ibid*, at 12–13. See also the general rule in American Jurisprudence Second (43 Am Jur 2d Insurance § 970 (1982)) which provides: 'A stockholder in a corporation has an insurable interest in the property of the corporation which will sustain a recovery on a policy issued to him on such property. Subject to some authority to the contrary, the rule is that a stockholder has an insurable interest in specific corporate property, although he does not possess an estate, either legal or equitable, in the property insured. Upon similar principle, the interest of a stockholder in the profits of an adventure undertaken by the corporation is also an insurable interest in such profits.' The California Insurance Code, s 281 provides that '[e]very interest in property, or any interest in relation thereto, or liability in respect thereof, of such a nature that a contemplated peril might directly damnify the insured, is an insurable interest.' Para 3401 of the New York Insurance Law, Art 34 defines insurable interest in property insurance as including 'any lawful or substantial economic interest in the safety or preservation of property from loss, destruction or pecuniary damage.' The Alaska Statute 21.42.030(a) defines insurable interest as 'an actual, lawful, and substantial economic interest in the safety or preservation of the subject of the insurance free from loss, destruction, or pecuniary damage or impairment.' See further the decision of the Alaska Supreme Court in *State Farm Insurance Company v Raymer* 977 P 2d 706 (1999).
[87] 161 US 57 (1896).
[88] *Ibid*, at 65. See further, (1982) 68 *Va L Rev* 640.
[89] 106 NY 535 (1887).

> that if there be ... a loss of the property [which] will cause pecuniary damage ... he has an insurable interest.[90]

This line of reasoning is further illustrated by *Castle Cars Inc v United States Fire Insurance Co*.[91] The issue was whether a *bona fide* purchaser for value, a car dealer, had an insurable interest in a stolen car which he had purchased for $2,600. Justice Poff, citing *dictum* in *Tilley v Connecticut Fire Insurance Co*,[92] noted that 'our Court has indicated that it considers the Eldon view too restrictive.' He rejected the insurer's contention that an insured must have a lawful interest in the subject-matter of the insurance. The judge reasoned that the dealer's interest was not acquired in violation of the law and that the word 'lawful' should not be viewed as synonymous with the word 'legal':

> An interest enforceable against the world is legal. An interest acquired in good faith, for value, and without notice of the invalidity of the transferor's title is lawful and enforceable against all the world except the legal owner. We share the view of those courts which have held that such an interest in a stolen vehicle is an insurable interest ... [93]

The innocent purchaser had paid for the stolen vehicle. He therefore had a continuing interest in its future existence and would suffer economic loss should it be damaged or destroyed. Yet another example of this broader approach to the definition of insurable interest is *Isabell v Aetna Insurance Co*.[94] Lester, a building contractor, was constructing a house for the plaintiffs, Mr and Mrs Isabell, under a 'turn-key' contract. Payment was, therefore, only to be made on completion and acceptance by the owner and the Farmers Home Administration (FHA), which was financing the project. FHA required the Isabells to insure the property during construction. The Isabells effected a fire insurance policy with Aetna, naming themselves as the insured and FHA as mortgagee; no mention was made of Lester. The policy contained a Dwelling Builder's Risk Endorsement. When the building works were nearing completion, the structure was destroyed by fire. Aetna repudiated liability on the basis that the Isabells had not sustained

[90] *Ibid*, at 540.
[91] 273 SE 2d 793 (1981). Although 'factual expectation' has not found favour in all US jurisdiction: see *Farmers' Mutual Insurance Co v New Holland Turnpike Co*, above, n 83 and, more recently, *Splish Splash Waterslides, Inc v Cherokee Insurance Co* 307 SE 2d 107 (1983). See also, the California Insurance Code § 283.
[92] 86 Va 811 (1890). Echoing Lawrence J's judgment, the court stated, at 813: 'Any person who has any interest in the property, legal or equitable, or who stands in such a relation thereto that its destruction would entail pecuniary loss upon him, has an insurable interest to the extent of his interest therein, or of the loss to which he is subjected by the casualty.'
[93] *Ibid*, at 814.
[94] 495 SW 2d 821 (1971).

any loss, and Lester was uninsured. Although the turn-key contract made no mention of insurance, Lester successfully sued the Isabells by showing there was an oral agreement that they would insure the structure for his benefit during its construction. The Isabells sued Aetna on the policy and the trial court found in their favour. Aetna appealed. The Tennessee Court of Appeals held that the defendant insurers were liable on the policy. In concluding that the Isabells had an insurable interest in the building, Carney J observed:

> Generally any person who derives a benefit from the existence of property or who would suffer loss from its destruction has an insurable interest therein and it is sufficient that loss of the property insured not only would, but might, subject the insured to pecuniary injury .... One having the care and custody or possession of property for another without liability and without any pecuniary interest therein may nevertheless obtain insurance thereon for the benefit of the owner.[95]

Agreeing with the trial judge, he went on to state:

> Many cases have held that a contractor who has agreed to construct, alter, or repair a building or other structure for a stipulated price has an insurable interest therein, whether the payment of the contract price is to be in instalments as the work progresses or upon completion of the job, at least where there is nothing in the contract exempting the contractor from the obligation to rebuild or complete the building or structure after damage or destruction. Likewise, persons furnishing materials for use in the construction of a building have an insurable interest therein.
>
> The owner of a building under construction has an insurable interest to the extent of its value, although the loss, in the absence of insurance, would fall on the contractor, and not the owner. This is because the title to the land carries with it the title to the building as completed. Even where a building is being erected by a contractor for its owner on land owned by a third person, the owner of the building has an interest in the building insurable under a builder's risk policy.[96]

The relative merits of the factual expectation and the legal interest tests were also demonstrated by *Royal Insurance Co v The Sisters of Presentation*.[97] An order of nuns had vacated an old convent for a new residence and had contracted to have the building demolished. During the course of demolition it was accidentally destroyed by fire. Notwithstanding that they had

[95] *Ibid*, at 825.
[96] Citing 43 Am Jur 2d, § 470, at 511.
[97] 430 F 2d 759 (1970). See also *Chicago Title & Trust Co v United States Fidelity and Guaranty Co* 376 F Supp 767 (ND Ill 1973).

retained title, the court held the order lacked pecuniary interest and therefore did not have the requisite insurable interest in the structure because the building was economically useless and had no value. Paradoxically, had the court employed Lord Eldon's test there would have been an insurable interest.

IV. INSURABLE INTEREST REVISITED BY UK COURTS

In the few cases where insurable interest is directly in issue,[98] it is not surprising that *Macaura* continues to represent the orthodox approach. For instance, in *Mitchell v Scottish Eagle Insurance Co Ltd*,[99] the pursuer, Mitchell, had entered into partnership with his son but had insured the partnership's premises in his own name. In the Outer House, Lord Prosser, applying the *Macaura* principle, held that Mitchell lacked an insurable interest.[100] More directly, in *Cowan v Jeffrey Associates*,[101] the issue again arose as to the interest possessed by the director and sole shareholder of a company. Lord Hamilton, sitting in the Outer House, felt obliged to follow *Macaura*, observing that, while it was an English authority and not, therefore, technically binding on him, nevertheless, it was highly persuasive. In his view the adoption of factual expectancy would require either legislative intervention or the House of Lords reversing itself.

Although these recent Scottish decisions show the continued importance of the narrow legal interest test, the judiciary has displayed tentative signs of a willingness to sidestep the force of *Macaura*. For instance, in *Sharp v Sphere Drake Insurance Ltd, 'The Moonacre'*,[102] Colman J distinguished *Macaura* and held that the sole shareholder in a company possessed an insurable interest in a yacht purchased by the company because the yacht was intended for his use and a power of attorney had been granted to him in respect of it.

It appears that fundamental corporate law doctrine was the driving force in *Macaura*.[103] In so far as proprietary interest determined the issue of insurable interest, the House of Lords necessarily concluded that this was vested in the company as a separate entity. An advantage of the factual expectation test, given its focus on economic interest, is that any insured who is able to demonstrate prejudice by the destruction of property may recover. If Macaura had, as is typical in close companies, depended on the business for his livelihood, then he would possess a pecuniary interest in the

[98] For the reasons why there have been so few cases, see, J Birds, 'Insurable Interests', in N Palmer and E McKendrick (eds), *Interests in Goods* (London: LLP, 1998).
[99] 1997 SLT 793.
[100] Lord Prosser drew support for his position from the judgment of Lord Sutherland in *Arif v Excess Insurance Group Ltd* 1987 SLT 473.
[101] 1999 SLT 757.
[102] [1992] 2 Lloyd's Rep 501.
[103] See *Salomon v A Salomon Co Ltd* [1897] AC 22.

timber on the basis of his factual expectation of benefit from its continued existence and prejudice from its destruction. He had a moral certainty of advantage from the company's asset (timber) but for the fire. Such an approach necessarily entails a determination of the type of company in question. In the context of small owner-managed companies, the wealth of modern jurisprudence surrounding statutory minority remedies shows that the courts are prepared to recognise that the nature of such companies requires a different approach from that which is taken towards large private or public companies. Moreover, this approach was not unknown at the time *Macuara* was decided.[104] Given the pecuniary proximity between shareholders in close companies and their assets it is also difficult to see why the moral hazard—the incentive to destroy the insured property—is any greater than if Macaura possessed a legal interest in the timber. By analogy the same argument can, of course, be applied to partnerships.

While the requirement of insurable interest has not been subjected to the sort of rigorous analysis that led other common law countries to adopt the factual expectation test, the courts have, nevertheless, taken cognisance of developments in those jurisdictions. The issue has emerged in the context of insurers' rights of subrogation. Simply put, on paying a claim in full the insurer takes on the rights of action which the insured would have had: in effect, the insurer steps into the shoes of the insured. This means that the insurer can, for instance, sue the tortfeasor responsible for the loss. However, the tortfeasor can use all the defences that would have been available against an action brought by the insured, and the insurer will have no action where the loss is caused by the insured.[105] This is of particular importance in insurance policies relating to construction contracts, in which the head contractor is commonly required by the contract to insure the project for their own benefit and that of the sub-contractors. The issue then comes down to whether the sub-contractor has an insurable interest which will grant immunity against the insurer's subrogation rights.[106]

In this context the courts distinguish between joint, composite and pervasive interests. While joint insurance covers joint owners of property (a typical example would be a policy on the matrimonial home taken out by spouses) and composite insurance provides cover for two or more persons interested severally, pervasive interest is a hybrid 'for the very good reason that in such a case the claimant is entitled to claim not only for himself but also for the benefit of his coassureds in the full amount of the loss.'[107]

[104] See, for example, the reasoning of Lord Cozens-Hardy MR in *Re Yenidje Tobacco Co Ltd* [1916] 2 Ch 426, and now *Ebrahimi v Westbourne Galleries Ltd* [1973] AC 360, Lord Wilberforce. See further P Davies, *Gower's Principles of Modern Company Law* (London: Sweet & Maxwell, 2003) Ch 27; and L Sealy, *Cases and Materials in Company Law* (London: Butterworth's, 2001), Ch 10.

[105] *Simpson v Thompson* (1877) 3 App Cas 279.

[106] Although for a critique see, J Birds, above n 98

[107] *State of the Netherlands v Youell and Hayward* [1997] 2 Lloyd's Rep 440 at 448, per Rix J.

This is illustrated by *Petrofina (UK) v Magnaload Ltd*,[108] in which it was held that the owners, the head contractors and each of the sub-contractors on a construction site had an insurable interest in the entire works despite the fact that they were working only on limited parts of the site. Their interest arose not from any ownership or possession, but from the fact that, in the event of negligence, each sub-contractor could suffer loss should any part of the works be damaged or destroyed. So, although the sub-contractors lacked property interest in the work in progress, they had an insurable interest in its continued existence. In so finding, Lloyd J drew the analogy of the insurable interest possessed by a bailee in goods.[109] Accordingly, it was possible for a policy covering the entire works to be taken out on a coinsurance basis by the head contractor and all sub-contractors. Lloyd J reasoned that to hold to the contrary would result in commercial inconvenience as each subcontractor would need 'to take out his own separate policy. This would mean, at the very least, extra paperwork; at worst it could lead to overlapping claims and cross claims in the event of an accident. Furthermore … the cost of insuring his liability might in the case of a small sub-contractor, be uneconomic.'[110] He, therefore, held that the sub-contractor's insurable interest lay in a 'pervasive interest' in the entire property. In the absence of any English decision directly on the point, Lloyd J turned to the decision of the Supreme Court of Canada in *Commonwealth Construction Co Ltd v Imperial Oil Ltd*, in which de Grandpre J stated:

> On any construction site, and especially when the building being erected is a complex chemical plant, there is ever present the possibility of damage by one tradesman to the property of another and to the construction as a whole … . By recognising in all tradesmen an insurable interest based on that very real possibility, which itself has its source in the contractual arrangements opening the doors of the job site to the tradesman, the Courts would apply to the construction field the principle expressed so long ago in the area of bailment. Thus all the parties whose joint efforts have one common goal, eg, the completion of the construction, would be spared the necessity of fighting between themselves should an accident occur involving the possible responsibility of one of them.[111]

[108] [1983] 2 Lloyd's Rep 91. Noted by J Birds [1983] *JBL* 497.
[109] *Tomlinson (Hauliers) Ltd v Hepburn* [1966] AC 451; *Waters & Steel v Monarch Fire and Life Assurance Co* (1856) 5 E & B 870.
[110] Above n 108 at 96–7. In a Canadian case that has been cited before the English courts, the issue has been reduced to one of construction of the particular terms of the insurance contract. In *G A Baert Construction (1960) Ltd v Canada Insurance Co* [1966] ILR 1–170 (Man CA), the insured was the owner and general contractor of a new building. The insurers issued a standard fire insurance policy, but added a Schedule, 'Builders Risk Completed Value Form', which extended insurance cover to 'All the property of the Insured or for which the Insured is legally responsible.' It was held that materials brought onto the construction site by subcontractors and the general contractor, and not yet incorporated into the building, constituted property for which the owner and the general contractor were legally liable, and were therefore within the terms of the policy. See also *Wilson v Jones*, above n 45.
[111] (1976) 69 DLR (3d) 558, at 561.

The approach taken by Lloyd J was endorsed in *National Oilwell Ltd v Davy Offshore (UK) Ltd*.[112] Colman J held that the suppliers of a subsea wellhead completion system for a floating oil production facility were coinsured's under the contractor's All Risks policy. The judge dismissed the contention that there could not be an insurable interest based merely on potential liability arising from the existence of a contract between the insured and the owner of property. Instead, he held that an insurable interest could be found in 'the insured's proximate physical relationship to the property in question.'[113] He went on:

> There is, in my judgment, in particular no reason in principle why such a supplier should not, and every commercial reason why he should, be able to insure against loss of or damage to property involved in the common project not owned by him and not in his possession. The argument that because he has no possessory or proprietary interest in the property he can have no insurable interest in it and that his potential liability in respect of loss of or damage to it is insufficient to found such an insurable interest is in my judgment misconceived.[114]

This expansive approach was explicitly adopted by the Court of Appeal in *Glengate-KG Properties Ltd v Norwich Union Fire Insurance Society Ltd*.[115] The court considered the meaning of the phrase 'the interest of the insured' contained in a policy covering the owner of a building against consequential loss following a fire or other insured peril. The issue was whether the insured could recover for the loss of architect's drawings, which were, at the time of the loss, owned by the architects, although the insured might eventually have acquired them. It was held that the insured had an insurable interest in the drawings despite their lack of proprietary interest. Although the court saw the insurable interest as being in respect of consequential loss rather than in the actual drawings themselves, Auld LJ and Sir Iain Glidewell expressed the view that the drawings could have been insured on the basis of Lawrence J's 'factual expectation' test.[116]

[112] [1993] 2 Lloyd's Rep 582.
[113] *Ibid*, 611. See also *Stone Vickers Ltd v Appledore Ferguson Shipbuilders Ltd* [1991] 2 Lloyd's Rep 288, 301, in which Colman J (sitting then as Mr Anthony Colman QC, a deputy judge of the High Court) stated that a risk of being materially adversely affected by loss of or damage to the contract works 'by reason of the incidence of any of the perils insured against' was capable of amounting to 'a sufficient [insurable] interest in the whole contract works' (reversed by the Court of Appeal on another ground without the issue of insurable interest being considered [1992] 2 Lloyd's Rep 578). Similarly, in *Hopewell Project Management Ltd v Ewbank Preece Ltd* [1998] 1 Lloyd's Rep 448 at 456, Mr Recorder Jackson QC, applying *National Oilwell UK Ltd*, stated that had the defendant contractors' satisfied the definition of 'insured' contained in the policy: 'the fact that the defendants were carrying out the professional services on site, and the fact that the defendants might incur legal liability for negligently causing damage to the contract works, gave the defendants an insurable interest in the contract works.'
[114] *Ibid*.
[115] [1996] 1 Lloyd's Rep 614.
[116] *Ibid*, at 623–4 and 625–6 respectively.

While not the primary issue in these cases, the definition of insurable interest was, nevertheless, decisive in determining the rights and obligations of the parties. This line of authority can, therefore, be seen as amounting to some recognition of the broader conception of interest as adopted in Canada and elsewhere. However, a limit was placed upon this trend by the Court of Appeal in *Deepak Fertilisers & Petrochemical Corporation v Davy McKee (London) Ltd and ICI Chemicals and Polymers Ltd*.[117] The court did accept the broad conception of insurable interest. It agreed that a sub-contractor in a building contract possessed an insurable interest in the entire works during construction. This was because of the economic disadvantage which would be suffered if, in the event of the structure being damaged or destroyed, they lost the opportunity to complete the work and receive remuneration. However, once the work had been completed the court stressed that such an interest came to an end.

The litigation arose out of the construction for Deepak of a methanol plant in India which exploded a few months after it was commissioned. There was a 'Marine-cum-Erection' policy under which contractors and sub-contractors were named as coassureds. The sub-contractors included Davy McKee, a firm of consulting engineers, and ICI, who had provided expertise and technology. The critical question was whether the coassureds had an insurable interest and were, therefore, immune from an action brought by Deepak's insurers exercising subrogation rights. At first instance, Rix J rejected the insurer's contention that the decisions in *Petrofina* and *National Oilwell Ltd* pointed to the cessation of insurable interest after completion of the works and concluded that Davy and ICI possessed the requisite interest so long as they were arguably responsible for damage to it. Since it was the insurer's case that Davy and ICI bore responsibility for the explosion even though it happened after completion of the plant, Rix J felt that they should 'in principle ... be entitled to insure against their potential liability.'[118] In his view the question relating to any temporal limitation on a contractor's insurable interest had not arisen in the earlier caselaw, but he saw no reason why an architect, technical designer, or constructor should not be able to insure against liability for damage due to their fault, even though it occurred after completion of the structure.

The Court of Appeal, reversing the judge's decision, held that the insurable interests of the coassureds ceased immediately upon completion of the works and did not continue thereafter merely because they could be exposed to potential liability at some future time. It followed that any damage to the plant after it was commissioned that was attributable to the fault of Davy could not be regarded as covered by the insurance policy.

[117] [1999] 1 Lloyd's Rep 387.
[118] [1998] 2 Lloyd's Rep 139, at 158.

The insurers, therefore, could sue Davy. Stuart-Smith LJ stated that once the works were complete:

> Davy would only suffer disadvantage if the damage to or destruction of the property or structure was the result of their breach of contract or duty of care. In order to protect the contractor and sub-contractors against the risk of disadvantage by reason of damage or destruction of the property or structure resulting from their breach of contract or duty, they would, in accordance with normal practice take out liability insurance, or, in the case of architects, professional indemnity insurance … . [W]hat they cannot do is persist in maintaining an insurance of the property or structure itself.[119]

Significantly, he did concede that the trial judge had been correct in stating that an architect, technical designer or contractor can participate in insurance under an All Risks policy because if the project were not completed due to the occurrence of an insured risk they would lose remuneration. Stuart-Smith LJ thereby adds further authority to the line of decisions beginning with *Petrofina*.[120] However, he took the view that an all risks policy on a building project is property insurance. As has been seen, this means that the project is insured against damage occurring during construction. The contractor and sub-contractors have insurable interests because they may suffer economic loss if building work ceases. Once the construction is completed, the contractors no longer have such an interest in the property and must effect a separate policy to cover any potential liability for their own negligence or breach of contract which results in damage to the building arising after completion. This maintains the distinction between property and liability insurance which, it has been argued, [121] was undermined by the decisions in *Stone Vickers* and *National Oilwell.*

The issue was revisited by the Court of Appeal in *Cooperative Retail Services Ltd v Taylor Young Partnership*.[122] The action arose as a consequence of a fire occurring on a construction site owned by the claimants, Cooperative Retail Services Ltd (CRS). The main contractor was Wimpey (W), the architects were Taylor Young Partnership (TYP), the electrical engineers were Hoare Lea & Partners (HLP), and the electrical sub-contractors, who worked on the building's generator system, were Hall Electrical (H). H entered into a warranty with CRS and W in which H undertook to exercise all reasonable skill. CRS claimed that the fire was caused by negligence

[119] Above n 117 at 399.
[120] Stuart-Smith LJ observed, '… Davy undoubtedly had an insurable interest in the plant under construction and on which they were working because they might lose the opportunity to do the work and to be remunerated for it if the property or structure were damaged or destroyed by any of the 'all risks', such as fire or flood', *ibid.* See J Birds and NJ Hird, *Birds' Modern Insurance Law* (London: Sweet & Maxwell, 2001) at 61.
[121] See J Birds, above n 98 and [2002] *JBL* 351.
[122] 74 *Con LR* 12.

or breach of contract on the part of TYP and HLP. TYP and HLP argued that the fire was caused by breaches of the main contract by W and breaches of warranty by H, and they, therefore, sought contribution from them under the Civil Liability (Contribution) Act 1978. Both the trial judge and Brooke LJ referred extensively to *National Oilwell* and *Petrofina*, with particular emphasis on Lloyd J's use of the *Commonwealth Construction* case. Brooke LJ accepted that CRS's insurers, acting through subrogation, could not pursue W or H because they were named as co-insureds under CRS's policy,[123] and, consequently, they were not liable to contribute to TYP and HLP because, as co-insureds, they were not 'other person[s] liable in respect of the same damage (whether jointly with him or otherwise)' under s 1(1) of the 1978 Act.

It is noteworthy that the broad conception of insurable interest has been utilised beyond the confines of construction insurance. As has been seen, *Stone Vickers*[124] concerned shipbuilding. In *Mark Rowlands Ltd v Berni Inns*,[125] the issue arose because of a tenancy agreement that required the lessor to insure premises against fire and to use any proceeds of insurance to rebuild. The tenant contributed to the premium and was relieved of the duty to repair in the event of fire damage. The Court of Appeal held that the insurers, who had paid out on the policy following a fire, could not recover against the tenant. The tenant was not mentioned in the policy, but it was clear from the terms of the lease that the insurance was effected on his behalf. The court took into account the fact that the tenant was required to contribute to the premium and that the lease excluded his liability for fire. Although the insurers were unaware of the arrangement in the lease, the court held that it must have been the intention of the lessor and lessee that in the event of a fire the lessor's loss would be recouped from the insurance policy. This meant there would be no other claim against the tenant by the lessor, or, therefore, by the lessor's insurer. In his judgment, Kerr LJ ignored Lord Eldon's definition and explicity adopted what he termed the 'classic' definition of insurable interest given by Lawrence J.[126]

[123] See *Simpson v Thompson* (1877) 3 App Cas 279.
[124] Above n 113.
[125] [1986] QB 211. Clearly, where the lessor and the tenant jointly take out an insurance policy, the insurer cannot use the covenant to keep the house in good repair as a means to sue the tenant for damage to the house because the tenant would be entitled to claim against the policy; unless, of course, the tenant has deliberately damaged the property, in which case he or she could not claim against the policy. The tenant will not be able to claim immunity from a subrogated claim by the lessor's insurers in respect of damage to any parts of the building which are not covered by the lease, as, for instance, where the tenant negligently sets fire to his or his part of the premises and that fire also damages premises occupied by another tenant of the same lessor in the same building: see *Barras v Hamilton* 1994 SLT 949. For another example in which an agreement with a third party excluded rights of subrogation see *Thomas & Co v Brown* (1899) 4 Com Cas 186. See also J Lowry and P Rawlings, *Insurance Law: Doctrines and Principles* (Oxford: Hart Publishing, 1999), at 205.
[126] Above n 124 at 228.

V. A CONTINUING ROLE FOR INSURABLE INTEREST?

The principles laid down in *Macaura* remain influential in the handful of cases directly concerned with insurable interest, such as *Mitchell* and *Cowan*. On the other hand, Coleman J felt able to avoid its application in *Sharp*, and the courts have been willing to develop a broader conception of insurable interest where the main issue concerned subrogation rights. The subrogation cases illustrate the need for modern insurance law to accommodate the complexity of modern commercial relationships. Of course, it is tempting to see these as separate issues: cases such as *Lucena* and *Macaura*, it might be argued, involve the validity of the insurance contract, while cases such as *Petrofina* and *National Oilwell* are concerned not with the validity of the contract, but rather with the proper determination of the parties covered. Leaving aside the confusion which might result from defining insurable interest in different ways according to the particular context, the distinction seems a little forced since in both cases there is a challenge to the rights of someone claiming to be insured under a particular policy. If, for instance, a sub-contractor is not a coinsured and can, therefore, be sued by the insurer, this must surely also mean that the sub-contractor can make no claims under the policy. The acceptance by the judges that sub-contractors can be coinsureds for the purpose of defending an action brought by the insurer seems equivalent to saying that the sub-contractor's loss is no less real than that suffered by the building's owner, even though they have no proprietary interest. This would suggest, in our view, that they should also be able to insure against such a loss. Finally, it might be suggested that the subrogation cases have arisen in particular contexts, most notably construction work, but this may simply be because of the size of the claims involved and because the complexity of the policy may raise real doubts as to whether the policy did provide cover.

The anxieties over moral hazard and wagering that prompted Lord Eldon to reach his view of insurable interest seem less relevant in the context of modern commercial practice. His test does not seem to achieve the objectives he believed it would. It does not necessarily provide any better deterrent against the moral hazard that the insured might destroy the property than the factual expectation test, indeed, it can be argued that an owner is likely to have more opportunity to damage the property.[127]

With respect to Lord Eldon's other concern, the dangers of a wager being concealed under the guise of an insurance contract seem more remote now than they were in the early nineteenth century. Since that time the activities of insurance and wagering have become separated. In spite of criticisms,

[127] See SA Rea Jnr, 'The Economics of Insurance Law' (1993) 13 *Int Rev of Law & Economics* 145–62. See generally, Malcolm Clarke, *Policies and Perceptions of Insurance: An Introduction to Insurance Law* (Oxford, Clarendon Press, 1997) at 21–22.

gaming has become a legitimate activity, as indicated by the Lottery and recent proposals to liberalise the gaming laws. As a consequence, the need to clothe wagering in the guise of insurance has been eliminated. In so far as wagering requires regulation, this is an issue of public policy and, therefore, rightly falls within the realms of public law. The use of insurable interest as a means of regulating this activity is inappropriate since, in general, it can only be raised at the whim of the insurers, and the factors influencing them to challenge a policy are likely to be matters of commercial interest, such as the impact it will have on future business, rather than matters of public interest. Indeed, it has been raised by insurers as a means of avoiding evidential burdens. *Macaura* itself has been explained as a case where the insurers believed that the insured had acted fraudulently, but, because they could not prove the point, they used insurable interest as a technical defence. Further, the statutory regulatory regime governing those who conduct insurance business reduces the likelihood of insurers concluding wagering contracts.[128] Finally, in spite of the 1745 Act and its successors, Policy Proof of Interest (ppi) contracts or honour policies, where the insurer agrees not to raise the issue of insurable interest, remain an important slice of the marine insurance industry.[129] While it would not make good business sense for the insurer to deny liability, even though such a policy is unenforceable, it seems curious that insurance law is so out of line with commercial practice.

A more fundamental point is that Lord Eldon prioritised the regulation of gaming through the use of insurable interest and failed to give proper emphasis to the competing public interest in ensuring that contracting parties perform their promises. In the final analysis, it is this that underpins the whole of contract law and might, therefore, be considered as of greater importance.

Modern insurers can frame coverage on the basis of a proposal in which they can ask such questions about the relationship between the proposer and the property as they think relevant to their decision as to whether or not to accept the risk. Moreover, the duty of disclosure, which places the insurer in an advantageous position when compared with parties in non-insurance contracts, makes it difficult to justify a situation in which the insurer can freely enter into the contract on the basis of full disclosure and still deny liability because of a lack of insurable interest. The implications of the lack of litigation directly on insurable interest since *Macaura* might be that the insurers accept the logic of this argument, or that the commercial implications of refusing to pay out would, presumably, be the same as those facing the bookmaker who failed to pay a winning bet.

Although some commentators have expressed the view that the courts are occasionally prepared to recognise that insurers may waive the requirement

[128] J Lowry and P Rawlings, above n 125, at 351 *et seq*.
[129] See above n 11.

of insurable interest,[130] it is our view that such a defence is not available to policyholders given that traditionally UK judges view the requirement as rooted in public policy.[131] In those cases where policies were enforced without insurable interest, such as *Prudential Staff Union v Hall*,[132] the courts recognised that on the facts the issue of wagering did not arise. It could also be argued that the courts are rightly ill disposed towards allowing lack of insurable interest to be pleaded where insurers are fully cognisant of the risk covered by the policy.[133] Recent support for this contention can be found in *Cepheus Shipping*, in which Mance J observed:

> … the present policy is not on its face one which the parties made for other than ordinary business reasons; it does not bear the hallmarks of wagering or the like. If underwriters make a contract in deliberate terms which covers their assured in respect of a specific situation, a Court is likely to hesitate before accepting a defence of lack of insurable interest.[134]

If the policy considerations that underpinned Lord Eldon's definition are no longer relevant, then one is left to wonder whether the requirement of insurable interest serves any useful purpose. It has the potential to allow the insurer to defeat the reasonable expectations of the parties and this encourages the judges to complicate the law by devising exceptions to the requirement as we have seen in the cases on subrogation. Where a party stands to suffer a pecuniary loss it seems illogical to argue that the desire to insure against such loss is tantamount to wagering. Parliament has acknowledged

[130] See J Birds, above n 105, at 94.

[131] In truth, English decisions have left open the question of whether an insurer can waive insurable interest, but on balance it would seem that such a defence is not available to insureds: see M Clarke, *The Law of Insurance Contracts* (London: LLP), para 4.1.D.

[132] [1947] KB 685. See also, *Thomas v National Farmers' Union Mutual Insurance Society* [1961] 1 WLR 386.

[133] In this respect, the position of insurers is safeguarded by the duty of disclosure: see J Lowry and P Rawlings, above n 125, at ch 4. See the judgment of Langley J in *Feasey v Sun Life Assurance Co of Canada* 2002 WL 819957, who, in considering the rationale underlying the Life Assurance Act 1774, observed that 'in the early 19th century the law of disclosure had not been developed to provide the protection to insurers which it does today.' See also Walter LJ's analysis in *Feesey Sun Life Assurance* [2003] EWCA Civ 885.

[134] *Cepheus Shipping Corporation v Guardian Royal Exchange Assurance plc (The 'Capricorn')* [1995] 1 Lloyd's Rep 622 at 641, per Mance J. Some hundred years earlier, Lord Brett MR had observed that, 'it is the duty of a Court always to lean in favour of an insurable interest, if possible, for it seems to me that after underwriters have received the premium, the objection that there was no insurable interest is often, as nearly as possible, a technical objection, and one which has no real merit, certainly not between the assured and the insurer', *Stock v Inglis* (1884) 12 QBD 564 at 571. In *Feasey v Sun Life Assurance Co of Canada, ibid,* Langley J also refers to this statement by Lord Brett MR. See further, *Mackenzie v Whitworth* (1875) 10 Exch 142 at 148, per Bramwell B; *Re London County Commercial Reinsurance Office Ltd* [1922] 2 ch 67 at 79, per Lawrence J; Australian Law Reform Commission, *Review of the Marine Insurance Act 1909,* above, n 78, ch 11. In different context, but to the same effect, see *Cleaver v Mutual Reserve Fund Life Association* [1892] 1 QB 147, 151, per

the difficulties caused by permitting third parties to sue insurers directly in certain circumstances, notwithstanding the absence of insurable interest.[135] One is left to wonder what it adds to the principle of indemnity under which, in general, the claimant is compensated for the pecuniary loss suffered. In *Macaura* the insurers were unable to substantiate their suspicion of fraud and this led them to raise the lack of insurable interest. As has been noted, the House of Lords also identified the difficulties inherent in assessing the loss suffered. It seems wrong to allow the requirement to be used as a technical defence in circumstances which bear no relation to its original policy objectives. Where fraud is alleged it should be proved. An insurer always has the option of refusing to underwrite a risk which is difficult to assess, such as where a shareholder seeks to insure the assets of a company.

The continuing insistence on requiring insurable interest—whatever definition is adopted—harks back to a time when policy issues dictated that this should be a precondition to the validity of the insurance contract. Once those policy arguments are removed, the justification for the requirement disappears.[136] Even if, as we have shown, a process of assimilation of the factual expectation test is underway, the obvious question remains, is there a role for insurable interest? It is our view that the principle of indemnity, which prohibits recovery to those who cannot establish proof of loss, is in itself sufficient to render the requirement redundant.

Lord Brett MR; *Saunders v Edwards* [1987] 1 WLR 1116, at 1134 (Bingham LJ). Generally, see J Lowry and P Rawlings, above n 125, at 159.

[135] See, for example, the Married Women's Property Act 1882, the Third Parties (Rights Against Insurers) Act 1930 and the Road Traffic Act 1988.

[136] On this see the comments of the Australian Law Reform Commission, above, *Insurance Contracts*, above, and *Review of the Marine Insurance Act 1909,* above n 78, para 11.24.

Commentary I on 'Re-thinking Insurable Interest'

SIR JONATHAN MANCE

JOHN LOWRY AND Philip Rawlings set out to demonstrate with clarity and force that a combination of two House of Lords decisions (*Lucena v Craufurd*[1] and *Mercaura v Northern Assurance Co Ltd.*[2]) and the Marine Insurance Act 1906 paints English law into the corner of an ancient field, but that this has not deterred English judges of a more modern generation from adopting an expanded view of insurable interest in adjacent fields. Adrian Hamilton [who provides the second commentary on this paper[3]], who was counsel in *Petrofina v Magnaload*,[4] ought, incidentally, to receive some old-fashioned prize money for his role in the latter context.

The rules governing insurable interest are certainly (a) technical, because of the various different statutory regimes, (b) doubtfully necessary in many cases, because of the principle that an insured can only recover an indemnity and (c) contrary also in spirit in other cases to the principle which allows valued policies covering sums considerably in excess of the actual loss.

Do the objections to the present position identified by Lowry and Rawlings matter? As they observe, it is not in insurers' insurance interests to raise too many points on insurable interest. Even in cases where they might be motivated to do so by their general view of the merits or lack of merits of a claim, insurers' general commercial interests may operate as some restraint. Attempts to insure by shareholders may not today be as frequent as in the past. But key-man or woman policies, taken out by companies in respect of their director or other leading spirit's lives or health, are a reverse image, yielding good premiums which insurers will not lightly imperil. The 1906 Act and the general principles of insurance law governing

[1] [1802] 2 B&P (NR) 269.
[2] [1925] AC 619.
[3] See below p 369.
[4] [1983] 2 Ll R 91.

disclosure and warranty offer examples of apparent inequity, which have been roundly criticised but impervious to reform. The doctrine of insurable interest is a more obscure problem, the practical implications of which are less obviously prejudicial.

And is the common law cause quite as lost as Lowry and Rawlings suggest? First, of course, the House of Lords can always override or qualify its previous decisions. But, secondly, it is worth taking a further look at existing authority. In *Moran Galloway & Co v Uzzielli*,[5] Walton J gave a judgment which the authors castigate as contradictory. But in reality he was affirming first that it was unnecessary to have a legal or equitable interest in the property insured, since second it also sufficed if there was a legal or equitable interest in some property that was 'dependant' on the safety of the thing exposed to the insured risks.

That, it seems to me, may leave some scope for development. Even s 5(2) of the Marine Insurance Act has some openness of texture, since it accepts that it is sufficient to stand in any legal or equitable relation to an adventure and not merely to any insurable property at risk therein.

It is relevant in this context to note *Wilson v Jones*.[6] It is part of the history of the great ship built by that next greatest Britain, Brunel, launched as the *Leviathan* and better known as the Great Eastern. Taken out of service as a passenger carrier within a few years of her maiden voyage, she was in 1865 chartered by the Atlantic Telegraph Company to lay the second transatlantic cable; the previous cable having been too weak and having broken after only three weeks (and 723 messages). The Great Eastern's first attempt in August 1865 ended in a broken cable and the litigation in *Wilson v Jones*. The breakage was graphically described in a letter from Sir Daniel Gooch, who was on board. He was a friend of Brunel and now owned the Great Eastern. To back the venture he had invested the huge sum of £¼ million in 'cable stock' (presumably issued by Atlantic Telegraph Company to finance the venture). Half the cable remained after the breakage at the bottom of the sea.

As to the litigation resulting from the breakage, the terms of the law report are both more prosaic and more obscure. The plaintiff was a very small shareholder in Atlantic Telegraph. He took out a marine policy in ordinary form, on the ship, and goods and merchandise for so much as concern the assured, valued at £200, say on twenty shares, valued at £10 per share. The Court of Exchequer Chamber upheld the claim, accepting that the plaintiff as shareholder could have no insurable interest in the cable, but describing the subject matter insured as being his interest in the adventure[7] or his interest in the laying of the cable on that particular voyage.[8]

[5] [1905] 2 KB 555.
[6] [1867] LR 2 Ex 139.
[7] *Ibid*, Willes J at p 149.
[8] *Ibid*, Blackburn J at p 152.

We move to a non-marine context, and what Lowry and Rawlings clearly regard as the dead-hand of *Mercaura v Northern Assurance Co Ltd*.[9] But it is to be noted that in that very authority, the House of Lords considered *Wilson v Jones*; and they did not reject or overrule it, but explained it as a case 'where the insurance was upon the adventure in which the shareholder had an interest, and not upon the cable in which he had none'[10] or was a case where although 'the policy described the subject-matter of the insurance in a very obscure manner, it was held that the shareholder insured had an interest that he could insure in the profits of the adventure so described, but it was expressly stated that he had no such interest in his shares in the company'.[11]

This shows two things at least. First, courts need not be too rigid in ascertaining the interest by reference to the literal language of the policy. A certain flexibility, to find and uphold an insurable interest, may be permissible. Secondly, that insurable interest may in some circumstances lie in the existence of an 'adventure' in which the insured is engaged and by which he aims to profit. If that is right, then very broad coverage can be sought and issued, since there are in law virtually no limits to the perils or risks that may be insured.

It has been suggested that the only admissible relationship to any such adventures is one which involves or arises out of some proprietary or contractual interest in it. So it is said that the shareholding in *Wilson v Jones* was critical. But that is hard to square with Lord Sumner's words. Further, if a rower sets out to row the Atlantic with the prospect of great financial reward if he succeeds, and he decides to insure the success of the adventure against storms, can it really matter whether he owns or has contracted for the use of his boat or has been lent it for free? Is the swimmer who sets out to swim across the Channel unable to insure against storms, because she uses no property? And what about general loss of profits insurances, as Clarke on Insurance Law points out? Must the concert pianist who wishes to cover his loss of profits confine himself to personal injury policies, or try to attach his interest in his profit-making abilities to his piano? Surely this cannot even represent the current common law, and *Wilson v Jones* suggests a chink of light in the otherwise impenetrable obstacles of prior authority.

Finally, you will be glad to hear that the Great Eastern, a notoriously unlucky vessel, did in 1866 return, this time successfully, to the fray. A cable was laid from Greenwich to Hearts Content, Newfoundland in June—July, and for good measure in August the Great Eastern was able to recover the lost half of the 1865 cable from the ocean bed. But this was only done after a new company, the Anglo American Telegraph Company had been formed, so that Mr Wilson still needed his success in the litigation in *Wilson v Jones*, and English law may still perhaps benefit by it too.

[9] [1925] AC 619.
[10] *Ibid*, Lord Buckmaster at p 628.
[11] *Ibid*, Lord Sumner, pp 630–1, with whom Lord Wrenbury agreed at p 633.

Commentary II on 'Re-thinking Insurable Interest'

ADRIAN HAMILTON, QC

THE REQUIREMENT OF insurable interest in the insured for a valid insurable policy is, like many principles of insurance, derived from historical considerations, arising from marine insurance. Back in 1745 anxieties existed that persons might insure property in which they had no interest, by way of a wager, rather than to obtain an indemnity against loss to that person's interest in the property. So the Marine Insurance Act of that year, reciting that the absence of interest had been productive of many pernicious practices, declared that policies made '*interest or no interest*' were null and void.

The authors review comprehensively the history of insurable interest since then, and consider the current position.The basic thesis is that there is no need for a requirement of insurable interest, because insurance gives an indemnity only against loss suffered by the insured. So if he has no interest, he will get nothing anyway. But, as the authors point out, the practice of issuing ppi policies subsisted, and the courts ignored the tell tale pinmarks where the ppi slips had been attached.

Also the authors point out that valued policies—and nearly all marine hull policies are valued policies—are not contracts of indemnity, in the strict sense. On the contrary such policies often provide for values much higher than the true value. This is perhaps underlined by *The 'Maira' (No 2)*,[1] where the House of Lords held that an agreed value of two or three times the true value was not illegal or unenforceable.

Insurable interest, however, remains a requirement, whatever its limits. In practice, however, I have found that proof of insurable interest has not caused major problems. In a number of insurance cases, particularly marine insurance cases, the insurers have put the claimants to proof of their insurable interest. This has almost invariably been by way of probing

[1] [1986] 2 Lloyd's Rep 12.

the genuineness of the claim, and no point on insurable interest has been pressed to trial. The insurers do not wish to discourage future business.

The authors trace some two hundred years of judicial decisions, starting with the Napoleonic wars, and the rights of the crews in enemy ships captured as prizes. True stories with the whiff of gunpowder about them! In *Lucena v Craufurd*,[2] the two classic rival tests were formulated by Lawrence J in his advice to the House (the factual expectation test) and by Lord Eldon LC in his speech (the legal interest test). During the nineteenth century support was given in the courts for both tests, but Sir Mackenzie Chalmers in the Marine Insurance Act 1906 adopted Lord Eldon's test in Section 5(2), by requiring a legal or equitable interest. The House of Lords in *Macaura v Northern Assurance*[3] (a non-marine case) came down comprehensively on Lord Eldon's side.

So the opportunity to accept the broader Lawrence test into English law, by judicial decision, was lost. This is unfortunate when one considers the extent to which commerce has been assisted over recent years by the increase in the range of risks insured which do not fit easily into the Eldon test, with its strict requirement of a legal or equitable relationship. The authors point out that in other common law jurisdictions like the USA and Canada the courts have substituted factual expectancy—the Lawrence test—in order that the reasonable expectations of the parties are fulfilled. Australia has adopted by statute the test of economic disadvantage for non-marine insurance, and the Australian Law Commission proposes the same test for marine insurance. In the UK, *Macaura* remains the law, but the judges, particularly the commercial judges, have shown their willingness to side step the effects of *Macaura*.

In *Petrofina v Magnaload*,[4] Lloyd J relied on a decision of the Supreme Court of Canada (*Commonwealth Construction v Imperial Oil*[5]) and on commercial convenience, to hold that each of the owners, contractors and sub-contractors had a '*pervasive interest*' in building works, to enable them all to be insured in respect of the works. The point at issue was whether the insurers could exercise rights of subrogation against a sub-contractor, who was a co-insured. They could not. At pp 96–7 Lloyd J dismisses the possibility of illegality for lack of a sufficient insurable interest.

Since then the tide has ebbed away from the strict Eldon test in the later decisions. Only the Court of Appeal decision in *Deepak v Davy McKee*[6] administered a slight check in holding that a sub-contractors' pervasive interest in the whole works ended once the work was completed. I would

[2] [1802] 2 B&P (NR) 269.
[3] [1925] AC 619.
[4] [1983] 2 Lloyd's Rep 91.
[5] [1976] 69 DLR (3d) 558.
[6] [1999] 1 Lloyd's Rep 387.

have preferred a decision that the pervasive interest continued while damage to property could involve the sub-contractors in legal liability. The other cases cited by the authors demonstrate the reluctance of judges to accept the implications of the Eldon test. Mance J in the *Cepheus* case[7] makes it clear that a court will hesitate before accepting a defence of lack of insurable interest, where an insurance contract deliberately covers the situation. Insurers should keep their promises.

So the authors conclude that the principle of indemnity is sufficient, without the need for insurable interest. But insurable interest is still required under English law, albeit that its scope is circumscribed. So it seems that only statutory reform will achieve the aims of the authors. In para 24 of the BILA report on Insurance Contract Law Reform, we point out that the technical rules of insurable interest have not in practice generally caused problems for claimants. We do however recommend statutory reforms on the basis of the Australian reforms—a test of economic disadvantage, rather than of legal or equitable interest.

[7] [1995] 1 Lloyd's Rep 622.

13

The Challenge of Modern Bankruptcy Policy: The Judicial Response

DAVID MILMAN

I. INTRODUCTION

BANKRUPTCY LAW[1] IS a product of legislative intrusion into the contractual/common law relationship subsisting between debtor and creditor. It represents a disruption of freedom of contract born out of necessity. That Parliamentary intervention can be traced back several centuries[2] and it is apparent for all to see today that modern bankruptcy law is dominated by statute. However, for us to concede that fact is not to denigrate the real input of the courts over the centuries;[3] the judiciary have always played a formative role in shaping bankruptcy legislation to modern social needs and in some cases that contribution has included the formation of entirely new legal principles. An example of the more limited former phenomenon can be seen in the way in which some courts[4] moulded the meaning of the concept of 'trader' (first used in bankruptcy legislation in 1571)

[1] This term will be liberally interpreted in this essay to encompass the treatment of personal insolvency outside bankruptcy, and in particular the IVA regime. Official bankruptcy policy is to encourage the use of bankruptcy alternatives to settle debt.

[2] The 1542 Act (34 and 35 Hen VIII c 4) is often credited with being the first general bankruptcy statute in English law. For the history of early bankruptcy legislation see Fletcher, *The Law of Insolvency* (Sweet & Maxwell, London, 3rd edn, 2002) at 1.012–8.

[3] Although the Bankruptcy Court proper only emerged in 1831 (1 and 2 Will IV, c 56) the courts had for centuries before that handed down important rulings affecting the evolution of bankruptcy law.

[4] On the haphazard development of the concept of the 'trader' by the courts see J Cohen, 'The History of Imprisonment for Debt and its Relation to the Development of Discharge in Bankruptcy' (1982) 3 *J of Legal History* 153 at 160, E Welbourne, 'Bankruptcy Before the Era of Victorian Reform' (1932–4) 4 *The Cambridge Historical Journal* 51 at 55 *et seq*. Some judges were keen to extend the boundaries of bankruptcy law; other judges preferred to leave debtors to the mercies of personal insolvency law. In reviewing the role of the courts it must be conceded that they received considerable assistance from the legislature in extending

to enable a wider range of debtors to gain access to the facility of bankruptcy as opposed to being left to the mercies of personal insolvency law where imprisonment for debt still flourished well into the nineteenth century. The 'common law of bankruptcy' can be found to have shown its hand in the development of the law of fraudulent preference by Lord Mansfield in the eighteenth century.[5] This concept, under which selective debt repayments could be avoided by the courts, was subsequently subsumed by legislation,[6] though the common law heritage within the preference mechanism can still assert itself, as the Privy Council found in *Lewis v Hyde*[7] where a lacuna in the statutory provisions was found to have been anticipated by the underlying common law. Other common law principles survive intact notwithstanding the dominance of statute—witness the pragmatic (and distributionally just) rule against double proof of bankruptcy debts,[8] the hotchpotch rule,[9] which is of increased importance in an age where cross border insolvency is becoming more prevalent, and the ethical demands placed upon trustees in bankruptcy by the rule in *Ex parte James*.[10]

Having established the historical credentials of the courts in developing the substantive principles governing the bankruptcy regime, we now turn to the subject matter of this paper, namely an examination of whether the courts still have a significant role to play today in shaping bankruptcy law. Are the judges merely implementing pre-determined statutory norms in a mechanistic fashion or is there continued scope for real judicial creativity? A related topic also forming part of our investigation is the extent to which the courts have shown commercial empathy or 'nous' in framing their decisions. In short, has the contribution of the courts been a constructive one from the viewpoint of insolvency practitioners operating within the field?

Before embarking on these investigations some general observations on the operational remit of the courts within the bankruptcy regime as a whole

the boundaries of tradership; indeed in some cases Parliament was compelled to intervene to counteract inconvenient judicial rulings. The distinction between traders and non-traders for bankruptcy purposes was abolished in 1861.

[5] *Worsley v De Mattos* (1758) 1 Burr 467, *Alderson v Temple* (1760) 4 Burr 2235, *Harman v Fisher* (1774) 1 Cowp 119, *Rust v Cooper* (1777) 2 Cowp 629.

[6] See Bankruptcy Act 1869 s 92. See now Insolvency Act 1986 s 340.

[7] [1998] 1 WLR 94. This case was specifically concerned with the relevant preference provisions under New Zealand law but there is no doubting the relevance of the case for English law. Lord Browne-Wilkinson justified the advice of the Privy Council on the basis that the statutory provisions had been constructed on the assumption that a preference in fact was required at common law—see *ibid*, at 99–100.

[8] *Re Oriental and Commercial Bank* (1871) 7 Ch App 99, *Deering v Governor of the Bank of Ireland* (1886) 12 App Cas 20. The rule against double proof received strong support from Neuberger J in *Re Glen Express Ltd* [2000] BPIR 456 where it was described as an 'overarching principle'.

[9] *Selkrig v Davis* (1814) 2 Rose 291, *Ex parte Wilson* (1871–72) LR 7 Ch App 490, *Banco de Portugal v Waddell* (1880) 5 App Cas 161. For a modern discussion see *Cleaver v Delta American Reinsurance* [2001] UKPC 6, [2001] 2 AC 328.

[10] (1874) 9 Ch App 609.

might prove instructive. Historically the courts' role has been **threefold**. Firstly, it controlled access to bankruptcy via the adjudication stage.[11] This input continues to this day even to the extent of self-generated bankruptcies initiated by debtors themselves.[12] The court then took on a supervisory purview as the bankruptcy progressed; initially this was very much a 'hands on' activity through the agency of the bankruptcy commissioners,[13] but in modern times that control element has been largely serviced through independent trustees in bankruptcy drawn from the private sector of insolvency practitioners.[14] Finally, the court had traditionally enjoyed the power to officially terminate bankruptcy, primarily through the medium of discharge.[15] It is in this area that the judicial function has been considerably attenuated with the advent in 1976 of automatic discharge from bankruptcy (ie, discharge that does not require a court order).[16] The judicial role with regard to delaying discharge in certain defined situations is all that is left of this emancipating jurisdiction.[17]

[11] The court has indicated that there is no unqualified right of access—deposit fees still have to be paid—*R v Lord Chancellor ex parte Lightfoot* [2000] QB 597 (a case involving a debtor's petition). Although the introduction of individual voluntary arrangements in 1985 has hived off as much as 25 per cent of all potential bankruptcy cases the courts do still play a considerable role in controlling access to that alternative way of dealing with personal debt. Thus there is residual discretion as to whether an interim order (formerly an essential prerequisite of the IVA model) is to be made—*Re Julie O'Sullivan* [2001] BPIR 534. Moreover the court cannot make available the IVA procedure in cases where Parliament has made no provision for it—see *Fletcher v Vooght* [2000] BPIR 435 (this case is no longer generally applicable as the Insolvency Act 2000 permits IVAs without a preliminary interim order—see s 3 and Schedule 3 para 7, which were brought into force on 1 January 2003).

[12] In *Re Coney* [1998] BPIR 333 the court had to decide whether a debtor qualified for bankruptcy on the grounds of insolvency. The debtor had been bankrupted on his own petition but had then had second thoughts about the merits of bankruptcy. A liquidity test was applied by the court to resolve this question and to hold that the bankruptcy order had indeed been validly made.

[13] The bankruptcy commissioners were appointed by the Lord Chancellor to oversee bankruptcy administration—for an explanation of their role see E Welbourne (above n 4). The current compromise model based on shared control is derived from the 1883 Act (often referred to as Chamberlain's Act).

[14] The public sector also has a role to play through the medium of the official receiver. Although the court's input is a more distant one, there is no doubting that it retains overall control of the proceedings as s 363 of the Insolvency Act 1986 stresses—on this see *Hardy v Buchler* [1997] BPIR 643. The ability of the court to review its own orders under s 375 (a form of internal judicial review) is also indicative of its power to oversee the proceedings.

[15] Bankruptcy could be brought to a conclusion through an order for annulment; the court's function in this respect is preserved—Insolvency Act 1986 s 282.

[16] See Insolvency Act 1976 ss 7 and 8.

[17] The courts do indeed enjoy the power under section 279(3) of the Insolvency Act 1986 to delay automatic discharge if there are matters requiring further investigation—see *Holmes v Official Receiver* [1996] BCC 246, *Hardy v Focus Insurance Co Ltd* [1997] BPIR 77, and *Jacobs v Official Receiver* [1999] 1 WLR 619. On discharge most debts are released and the courts have been alert not to allow the exceptional category of surviving debts to expand as this would run counter to official policy—see *Mander v Evans* [2001] 1 WLR 2378, *Woodland-Ferrari v UCL Group Retirement Benefits Scheme*, [2002] EWHC 1354 (Ch), [2002] 3 WLR 1154.

II. FOUNDATIONS OF MODERN BANKRUPTCY LAW

The current legislative regime governing bankruptcy and personal insolvency law is to be found in the Insolvency Act 1986, a substantial piece of consolidating legislation that replaced the largely unconsummated 1985 Act of the same name. That legislative structure (the fundamentals of which had until recently remained largely unamended in the past two decades[18]) owed much to the constructive reforms recommended by the Cork Committee[19] which reported in 1982. Looking at the output of that body it is possible to identify some dominant insights and strategies that have informed the modern statutory matrix.

As far as the *consumers* of bankruptcy law were concerned, considerable emphasis was placed on liberalising the law from the perspective of the debtor. Cork acknowledged the fact of life of credit in a modern consumer society and took the view that, having accepted the economic benefits that flow from the institution of credit, society ought to operate a credit default mechanism such as bankruptcy in such a way as to recognise that debtors play a role in our prosperity.[20] Having taken this position, Cork recommended[21] deregulation via the removal of the automatic requirement for public examination, through shorter automatic discharge periods and by a strategy designed to repay debts through the debtor's income rather than by selling up assets in every case. Linked to this liberalisation was a conviction that the law should encourage the rehabilitation of debtors rather than grinding them down to become a drain on the social welfare system for the future. Finally, there was appropriate recognition given of the fact that debtors often exist within family structures and the law should take greater account of the needs of family members when determining the respective rights of a bankrupt and creditors. A current observer would find such policies unremarkable, but what has to be recognised is that at the time they were floated they ran counter to established judicial conventions which were very much inclined towards facilitating creditors. The tradition in English law of favouring creditors over debtors is well known.

Since the momentous statutory reforms of the nineteenth century, bankruptcy in English law has been characterised as an *institution of stewardship*.

[18] Thus there had been some tinkering at the technical margins—Insolvency (No 2) Act 1994 (transactional avoidance), Appointment of Trustees and Trusts of Land Act 1996 (family home) and Welfare Reform and Pensions Act 1999 (pensions). More significant reforms were enacted to the individual voluntary arrangement procedure by the Insolvency Act 2000 but these only took effect at the start of 2003. Major bankruptcy reform is to take effect in 2004 when the personal insolvency provisions in the Enterprise Act 2002 are brought into force.

[19] Cmnd 8558.

[20] Cmnd 8558 para 23.

[21] *Ibid*, para 599 *et seq*.

Creditors no longer enjoy the day-to-day control of the bankrupt's affairs that they once exercised.[22] The bankrupt is no longer in thrall to his creditors. The Cork Committee felt that the management of bankruptcies could be improved in several ways, most notably through the introduction of compulsory professional qualification for insolvency practitioners. Moreover, trustees in bankruptcy had to be given adequate powers to trace and recover assets so as to maximise the pool of assets available for distribution amongst creditors. The institution of stewardship within bankruptcy law is well regulated today. One comparative issue we can investigate is the level of regulation of stewardship in the context of the individual voluntary arrangement.

Bankruptcy is at its heart a *legal procedure* for the settlement of debt. As a procedure no one looking at its elements twenty years ago would contest the fact that it required an upgrade. It needed to move on from procedures set in place nearly a century earlier. Accordingly, Cork made recommendations designed to speed up the processes of initiation and management of bankruptcy.[23] Old anomalies, pregnant with judicial rulings over the ages,[24] were swept away. Again these reforms required a transformation in contemporary judicial culture.

The crown jewels of any bankruptcy are the *assets* forming part of the estate. Again the Cork Committee felt that there was scope for improving the rules defining the estate parameters and in particular in ensuring that the value of the estate was maximised in a manner that was consistent with modern expectations of fair play. The curious (and increasingly symbolic) doctrine of reputed ownership was consigned to the history books.[25]

Finally, and perhaps most importantly, Cork recognised[26] that bankruptcy should be seen as an *institution of last resort*. Although it was undoubtedly an advance on imprisonment for debt, it was no panacea. In particular, new mechanisms should be developed to offer alternative (and perhaps more efficient) ways of dealing with debt. This displacement strategy has borne fruit with the genesis (and undoubted success) of the individual voluntary arrangement (IVA) procedure.

[22] See M Lester, *Victorian Insolvency* (Oxford, Clarendon Press, 1995) chs 4 and 5 for the history of this change. Essentially Parliament permitted creditors to take control of the estate under the terms of the Bankruptcy Act 1869 but this experiment failed and a return to officialism (ie, control by a public official) was effected in the Bankruptcy Act 1883.

[23] Cmnd 8558 ch 11.

[24] Such as the voluminous case law on what constituted an act of bankruptcy (of which there were more than a dozen)—on this concept see Bankruptcy Act 1914 s 1(1).

[25] The doctrine dates back to s 11 the Bankruptcy Act 1623 (1 Jac I, c 19). Initially, the courts seemed happy to embrace (and even extend) the doctrine—*Ryall v Rolle* (1749) 1 Atk 165. However, hostility to it grew as it became obvious that it made a nonsense of modern commercial practices linked to credit—see *Belcher v Bellamy* (1848) 2 Exch 303 per Parke B and *ex parte Wingfield* (1879) 10 Ch D 591 at 594 per James LJ for critical judicial comment on the doctrine.

[26] See for example Cmnd 8558 para 549 where the view was expressed that bankruptcy was to be reserved for the more serious cases of personal insolvency.

In the following discussion these several insights into modern bankruptcy policy must be borne in mind. Have the courts adopted these as aids to judicial decision-making? Are these goals always internally consistent? What is the likely outcome if competing goals come into conflict?

III. EXTRINSIC TRENDS AND FACTORS RELEVANT IN BANKRUPTCY LITIGATION

A further complication facing the courts in adjudicating contested bankruptcy cases must now be noted. In the two decades since the modern structures were put in place new developments, whether they be social, economic or legal have exacerbated the challenge faced by the courts in processing bankruptcy litigation.

For instance, the increased official emphasis on encouraging individuals to take adequate steps to provide for a pension for themselves in retirement has brought bankruptcy law into conflict with governmental social policy encouraging self-provision for retirement. After much litigation,[27] amending legislation[28] has had to be introduced to effect a compromise and thereby to neutralise judicial decisions, which were entirely consistent with the new bankruptcy policy goals but which nevertheless proved to be socially inconvenient by undermining incentives to make pension provision for oneself.

The fundamental reforms to civil procedure implemented in the wake of the Woolf Report[29] might be expected to have had some impact on bankruptcy procedures. Fortunately, these reforms are consistent with the new procedural approach recommended by Cork and have caused fewer difficulties for the courts.

A more significant background change was the enactment of the Human Rights Act 1998.[30] Although the Cork approach favouring more liberal treatment of bankrupts had made the lot of such individuals much more agreeable than previously was the case, there were still a number of situations where it could be argued that their treatment was inconsistent with modern

[27] *Re Landau* [1998] Ch 223, *Krasner v Dennison (Lawrence v Lesser)* [2001] Ch 76, *Rowe v Sanders* [2002] EWCA Civ 242, [2002] BPIR 847.

[28] Pensions Act 1995 as modified by the Welfare Reform and Pensions Act 1999, s 11 which reverses *Re Landau* (above).

[29] Access to Justice: Final Report, (HMSO, London 1996). Although the CPR as such do not formally apply to insolvency proceedings the effect of the revised version of IR 1986 r 7.51 is to achieve that very end unless the Insolvency Rules provide otherwise in an individual case. For a review of the effect of the CPR on insolvency procedures see G Davis, 'Insolvency Proceedings in the Age of Woolf' [2000] *Ins Law* 33.

[30] This legislation took effect on 2 October 2000. For discussion see CA Gearty, 'Insolvency and .... Human Rights?' [2000] *Ins Law* 68, N Pike, 'The Human Rights Act 1998 and its Impact on Insolvency Practitioners' [2001] *Ins Law* 25, and W Trower, 'Human Rights: Art 6—The Reality and the Myth' [2001] *Ins Law* 48.

perceptions of fundamental rights. The courts have had to review settled principles of bankruptcy law in the light of this new rights-based climate.

Finally, the evolution of matrimonial law, and in particular the now more frequently resorted to rules governing financial settlement on divorce, have in the past twenty years thrown up a number of contradictions with established bankruptcy principles. Real problems of legal interface have been encountered and major perturbations have occurred.

IV. GENERAL APPROACH TO INTERPRETATION OF THE 1986 LEGISLATION

The courts were quick to recognise that the reforms in 1985–6 were not to be viewed as examples of legislative tinkering with the old established law of bankruptcy but rather to be construed as a completely fresh start. This is apparent from a number of early judicial pronouncements. Thus in *Re A Debtor (No 1 of 1987)*[31] Nicholls LJ, in dealing with an issue relating to a defective statutory demand (a topic we review later), declared:

> I do not think that on this the new bankruptcy code simply incorporates and adopts the same approach as the old code. The new code has made many changes in the law of bankruptcy, and the court's task, with regard to the new code, must be to construe the new statutory provisions in accordance with the ordinary canons of construction, unfettered by previous authorities.

In the context of the protective effect of bankruptcy in *Smith v Braintree District Council*[32] Lord Jauncey elaborated:

> … the Act of 1986, although re-enacting many provisions from earlier statutes, contains a good deal of fresh material derived from the Insolvency Act 1985. in particular, the legislation now emphasises the importance of the rehabilitation of the individual insolvent, it provides for automatic discharge from bankruptcy in many cases, and it abolishes mandatory public examination as well as enabling a bankrupt to be discharged without public examination. Thus not only has the legislative approach to individual bankruptcy altered since the mid—19th century, but social views as to what conduct involves delinquency, as to punishment and as to the desirability of imprisonment have drastically changed ….. in these circumstances, I feel justified in construing section 285 of the Act of 1986 as a piece of new legislation without regard to 19th century authorities or similar provisions of repealed Bankruptcy Acts.

[31] [1989] 1 WLR 271 at 276.
[32] [1990] 2 AC 215 at 237–8.

Since these initial heady days a degree of reality has intruded into the processes of judicial interpretation. Yes, the 1986 legislation did embody many new ideas, but it also contained much that was well established and the subject of a well developed body of case law. It would be foolhardy for the courts to abandon decades of informed jurisprudence. Accordingly, in *Re A Debtor (No 784 of 1991)*[33] Hoffmann J explained:

> Those authorities show that, in approaching the language of the Act of 1986, one must pay particular attention to the purposes and policies of its own provisions and be wary of simply carrying over uncritically meanings which had been given to similar words in the earlier Act. It does not, however, mean that the language of the new Act comes to one entirely free of any of the intellectual freight which was carried by words and phrases in earlier bankruptcy or other legislation.

Commenting on this cautionary note in *Cadbury Schweppes plc v Somji*[34] Robert Walker LJ elaborated:

> the intellectual freight least likely to be jettisoned includes the basic doctrines (such as proportionate treatment of unsecured creditors, and the principle of set-off) which have been features of English bankruptcy law since its earliest days. Although the English law of bankruptcy now has the appearance of a complete statutory code, it is built on foundations which owe much to past judicial creativity … .

Thus, the courts have empowered themselves to disregard apparently binding precedents where they wish or to apply them in other circumstances. In other words, they have reserved for themselves maximum flexibility in the process of judicial decision making. In a climate where bankruptcy policies are changing and other legally backed socio-economic policies have emerged to the point of conflict this ability to pick and choose precedents has proved most valuable. There is evidence to be cited of old authorities being discarded.[35] Equally there are many signs of the desire to promote the continuity of certain bankruptcy principles.[36] No consistent or predictable pattern can be detected.

[33] [1992] Ch 554 at 558–9. See also the approach of Ferris J in *Re Landau* [1998] Ch 223.

[34] [2001] 1 WLR 615. This case deals with individual voluntary arrangements but it shows the intimate connection with bankruptcy jurisprudence.

[35] *Smith v Braintree DC* [1989] 3 WLR 1317 overruling *Re Edgcome* [1902] 2 KB 403. The disclaimer case of *Hindcastle v Attenborough Estate* [1997] AC 70 is also another good example of the radical approach. Here in dealing with the effect of disclaimer on guarantors of leases the House of Lords overturned *Stacey v Hill* [1901] 1 KB 660 notwithstanding its well recognised status as a major authority in the area. Where there is a subtle change in the wording of the legislation that may persuade the court to depart from previous authority on counterpart provisions—on this see *Port v Auger* [1994] 1 WLR 862 at 873–4.

[36] The limited case law on s 303 might be taken as pointing to this conclusion. See also the limitation case of *Anglo Manx Group v Aitken* [2002] BPIR 215 where the court declined an invitation to disregard the authority of *Re Benzon* [1914] 2 Ch 68.

V. INITIATING BANKRUPTCY

The 1986 Act on the recommendation of the Cork Committee sought to simplify the bankruptcy procedure by removing the archaic (and convoluted) procession from acts of bankruptcy, through receiving orders up to the moment of adjudication. The most common of the many potential acts of bankruptcy consisted of failure to comply with a bankruptcy notice issued by a court. Instead, in future, the procedure was to be based exclusively[37] upon a statutory demand served outside court which, if unsatisfied, would lead to a petition and ultimately a judicial determination on whether a bankruptcy order should be granted.[38] This change of emphasis reduced the judicial input in initiating bankruptcy to one of final adjudication.

Under the former bankruptcy regime any error by the creditor in undertaking the bankruptcy procedure would prove fatal to his case and would necessitate a return to stage one; hardly a model for efficient case management. The rationale behind this procedural variation on 'snakes and ladders' was that as bankruptcy was a quasi-penal regime the responsibility of the court was to offer the maximum protection to debtors against any procedural irregularity.[39] Typical of this reasoning is *Re A Debtor (No 21 of 1950)* where Harman J frowned upon errors in the bankruptcy notice which referred incorrectly to the issuing court.[40]

That pernickety approach quickly fell victim to a new judicial realism in the wake of the 1985–6 reforms.[41] The watershed here was to be found in the Court of Appeal's ruling in *Re A Debtor (No 1 of 1987)*.[42] Here a statutory demand had been served using the wrong Form and containing incorrect information as to the amount of the alleged debt. The Court of Appeal held that as there was no dispute as to the fact that a sum in excess of the minimum bankruptcy level was due there would be no injustice to

[37] Compare the significance of statutory demands in winding up cases—*Re A Debtor (No 544/SD/98)* [2001] 1 BCLC 103.

[38] This is a matter for judicial discretion, though that discretion is subject to qualifications imposed by s 271. The courts also play a key part in weeding out weak and premature cases by setting aside statutory demands.

[39] Similar protective sentiments underpinned the ruling in *Fowler v Padget* (1798) 7 Term Rep 509 which required proof of mens rea before an act of bankruptcy could be deemed to have been committed. For the significance of this ruling see MS Servian, 'On the Demise of Acts of Bankruptcy' (1988) 4 IL & P 117.

[40] [1951] 1 Ch 313. See also *Re A Debtor (No 75 of 1982)* [1984] 1 WLR 353.

[41] Cork was critical of the former approach—Cmnd 8558 para 118. For discussion of this change in judicial attitudes see D Milman, 'Statutory Demands in the Courts: A Retreat from Formalism in Bankruptcy Law' [1994] Conv 289.

[42] [1989] 1 WLR 271. For a similar relaxed attitude see *Cartwright v Staffs and Moorlands DC* [1998] BPIR 328 (wrong Form used), *Morley v IRC* [1996] BPIR 452 (debt overstated), *Khan v Breezevale* [1996] BPIR 190 (security not stated) and *Re Blackman* [1999] BCC 446 (failure to state consideration for debt).

the debtor in letting the case proceed. This new way of looking at matters may well reflect the fact that bankruptcy is not the trauma for debtors it once used to be. More significantly, these cases can be explained on the grounds that the courts are opposed to the waste of judicial time and to the incurring of additional costs inherent in a 'return to square one' decision. The courts have therefore facilitated the use of the bankruptcy process by creditors and have removed potential traps from the way of practitioners. Having said that, Nicholls LJ pointedly warned banks and other creditors not to become too lax in their approach towards statutory demands.[43]

Although these decisions do reflect a more relaxed judicial perspective towards procedural shortfall in bankruptcy[44] that latitude has its boundaries. This is apparent from *Re Awan*,[45] where there was an error as to service of the bankruptcy petition. This was seen by Judge Boggis QC as a fundamental flaw which the courts did not feel justified in turning a blind eye to. The statutory relief made available by Insolvency Rules 1986 r 7.55 for waiving procedural departures was not intended to compensate for such a major error.

The new attitude of the court extends beyond procedural glitches. For example, it now seems happy to disregard the rule in *Ladd v Marshall*[46] and hold that fresh evidence could be admitted at the hearing of a bankruptcy petition, even though the issues in dispute had been aired at the earlier set aside application.[47] The point being here that the hearing of the set aside application is not viewed as a full trial of the issues, therefore the *Ladd v Marshall*[48] principle can be discarded without the principle of finality of adjudication being compromised. Clearly the same cannot be said of the petition hearing which is regarded as a full trial thereby precluding the admission of fresh evidence on an appeal.[49] A relaxed view of timing matters is also apparent from a perusal of personal insolvency legislation and its application by the courts.[50]

[43] [1989] 1 WLR 271 at 280.

[44] In IVA cases the courts have been more ambivalent towards laxity—compare *Mytre Investments v Reynolds (No 2)* [1996] BPIR 464 (where a strict approach was taken) with *IRC v Duce* [1999] BPIR 189 where procedural errors were not treated as fatal and creditors were given a second chance to vote on an IVA proposal. This issue is discussed further below.

[45] [2000] BPIR 241.

[46] [1954] 1 WLR 1489.

[47] *Royal Bank of Scotland v Binnell* [1996] BPIR 352, *Salvidge v Hussein* [1999] BPIR 410.

[48] Above.

[49] *Lombard Natwest Factors Ltd v Arbis* [2000] BPIR 79.

[50] *Tager v Westpac Banking* [1997] BPIR 543 applying s 376 of the Insolvency Act 1986. For further consideration of principles governing judicial discretion in such cases see *Warley Continental Services Ltd v Johal*, *The Times* 28 October 2002. Compare *Re Bournemouth and Boscombe AFC Ltd* [1998] BPIR 183 where time limits were held to be absolute in a CVA case because there was no statutory authority to waive time limits in matters of corporate insolvency.

In dealing with pre bankruptcy steps the courts have shown an awareness of commercial developments in general and have been prepared to facilitate change. In *Re A Debtor (No 51/SD/1991)*[51] a statutory demand citing a debt in a foreign currency was held by Morritt J to be acceptable as there was no specific prohibition in the legislation against framing a statutory demand in this manner. In an era of increased prevalence in cross border commerce this is entirely pragmatic.

At the end of the day there is discretion vested in the courts by section 266(3) of the Insolvency Act 1986 to dismiss or stay the petition and that discretion is there to be used as the Court of Appeal forcefully asserted in *Re Ross*.[52] Having said that, critics have argued that the judicial role on initiation is still largely mechanistic. Thus, if the debt is undisputed the order will normally be made.[53] Equally, the courts will not refuse to make an order simply because the debtor has made an offer of settlement to the creditor unless it is of the opinion that the creditor is acting unreasonably.[54] Here we see evidence of continued support for the position of creditors and also an understanding of the fact of life that last minute offers from debtors to settle their obligations rarely produce the promised goods.

VI. THE ESTATE AND OTHER DEBTOR RESOURCES

The assets of the bankrupt provide the battleground over which most contested bankruptcy litigation is fought. Those assets provide the possibility of debt recovery for creditors; but they may also offer the hope of redemption for a debtor seeking to embark upon a fresh life. A conflict is thus created and it is for the courts to resolve that tension in each individual case.

The general approach of the courts, aided and abetted by the broad statutory view of 'property' as indicated in section 436 of the Insolvency Act 1986, has been to include as much of value as possible in the estate so as to be available for distribution to creditors.[55] The courts have shown themselves as adept in recognising many commercially valuable rights as

[51] [1992] 1 WLR 1294.
[52] [2000] BPIR 636.
[53] It is sometimes said that a creditor is entitled to the bankruptcy order *ex debito justitiae* but that entitlement is not absolute. Having said that, the courts are reluctant to treat a petition by a genuine creditor as an abuse of process even if the creditor may have mixed motives in seeking the bankruptcy of the debtor—see *Hicks v Gulliver* [2002] BPIR 518.
[54] It has power to determine such questions under s 271 of the Insolvency Act 1986 but it will only intervene if it believes that the creditor is acting in a manner in which no reasonable creditor would act—see *Re Gilmartin* [1989] 1 WLR 513, *O'Brien v IRC* [2000] BPIR 306. Thus a test similar to the *Wednesbury* test in administrative law is applied and the fact that this test is used shows a strong inclination towards the position of the creditors.
[55] *Dear v Reeves* [2001] EWCA Civ 277, [2002] Ch 1—here a conveyancing authority on a related point was disregarded. See also *Re Rae* [1995] BCC 102 and *Rothschild v Bell* [2000] QB 33. Compare *Re Campbell (A Debtor No 145/1995)* [1996] BPIR 238.

property for the purposes of inclusion within the bankruptcy estate. They have clearly moved with the times.

In *Cork v Rawlins*[56] the question before the Court of Appeal was whether a sum of money representing an insurance payout for disablement could be claimed by the trustee as part of the estate or reserved for the future upkeep of the bankrupt. The problem facing the court was the well established common law principle[57] excluding compensation for personal injury from the ambit of the estate. Whilst acknowledging that this was a difficult case, the Court of Appeal gave primacy to the interests of the bankrupt's creditors over the optimistic expectations of the bankrupt. This was done by restricting the exception to compensation representing relief for pain and suffering and not extending it to compensation for loss of earning capacity. Chadwick LJ concluded his judgment in the following terms:

> It is plainly in the general public interest that persons should be encouraged to make provision against the possibility that they will be unable to meet their commitments as a result of misfortune for which they are not responsible. But if public policy requires that they should be encouraged to do so by permitting them to shelter that provision from the claims of their creditors, then it is for Parliament to say so. It is not, in my view, for the courts to distort the bankruptcy code in order to achieve that result.[58]

A similar philosophy has manifested itself with regard to pension benefits.[59] For many middle class bankrupts, pension rights, apart from ownership of a home, represent the major fruit of years of industry. Formerly, in the years prior to 1997, it was assumed that these would not become vested in the trustee, but conversely it was accepted that the income might become liable to be attached for the benefit of creditors. These comfortable assumptions were shattered by Ferris J with his controversial ruling in *Re Landau*[60]

[56] [2001] EWCA Civ 202, [2001] Ch 792.
[57] *Beckham v Drake* (1849) 2 HLC 579. The inevitable problem of a hybrid claim (compensation for personal injury and loss of earnings) arose in *Ord v Upton* [2000] Ch 352 and was solved by vesting the entire claim in the trustee in bankruptcy though with the proceeds of claim with respect to personal injury being held exclusively for the benefit of the bankrupt. See R S Sharpe, '*Ord v Upton*: Some Recent Developments' [2001] *Ins Law* 182.
[58] [2001] EWCA Civ 202 at para 39, [2001] Ch 792 at 803.
[59] For discussion in general see I Greenstreet, 'When Can a Trustee in Bankruptcy Get His Hands on a Pension: The post-Pensions Act 1995 Position' (1995) 11 IL & P 168, M Simmons, 'How Safe are Pensions in Bankruptcy?' (1997) 13 IL & P 98, A Deacock and A Martin, 'The Right of a Trustee in Bankruptcy to the Bankrupt's Pension: Pensions Industry *v* Insolvency Practitioners—A Score Draw' (2000) 16 IL & P 127, I Greenstreet, 'Pensions and Bankruptcy—Recent Developments' (2001) 17 IL & P 43.
[60] [1998] Ch 223—a critical element of the judgment was the refusal to uphold the old authority of *Re Huggins* (1882) 21 Ch D 85 (which treated pensions as income caught by the income payments order regime) on the grounds that the 1986 legislation had to be construed de novo unburdened by the intellectual freight of history. At the time it was handed down this judgment of Ferris J provoked hostility from the pensions industry and spawned many instant seminars for practitioners. For comment on the case see I Greenstreet, 'Practical Steps to Realise Value from a Bankrupt's Personal Pension Following *Re Landau* ' (1997) 13 IL & P 101.

where it was held that personal pension benefits vested automatically in the trustee. In spite of howls of outrage this judgment has stood up well and has been approved by the Court of Appeal,[61] most recently in *Rowe v Sanders*.[62] Legislation has been required to rectify the position and to restore some semblance of a status quo.[63]

The willingness of the courts to adopt an inclusive view of the modern bankruptcy estate has its limitations. In the well known case of *Haig v Aitken*,[64] where a trustee in bankruptcy sought to claim for the estate the private correspondence of a disgraced political figure, an outraged Rattee J declared:

> In my judgment it is inconceivable that Parliament really envisaged, by passing the Insolvency Act, that the effect of bankruptcy should be that a bankrupt's personal correspondence should be available for publication to the world at large by sale at the behest of the trustee in bankruptcy.

This is an interesting ruling that bucks the trend of authorities in a number of ways. The courts have been most adept at including within the estate any asset having a commercial value. There is no doubt that the documents in *Haig v Aitken*[65] did have some commercial worth, though the precise magnitude was a matter of dispute. It is difficult to rationalise this case in the light of surrounding case law and it is best seen as a one-off.

One of the more difficult areas to be traversed by the judiciary has concerned the competing rights to the family home on bankruptcy. The problem here being that this usually represents the major asset of the bankrupt, but by its very nature it will be subject to claims (whether they be legal or moral) from the rest of the family. Although the 1986 legislation as amended[66] did indicate a swing in the balance of power away from the strict legal rights of creditors, in many cases those rights were merely being deferred rather than extinguished. Further complications have arisen because of the requirements of the Human Rights Act 1998 and their potential impact in this field. In *Mountney v Treharne*[67] it was held by Stanley Burnton J that the right to family life specified in Article 8 ECHR did not give an *absolute* guarantee to a spouse or former spouse that the family home could not be claimed by the trustee. Although the ruling in this case was overturned on appeal the *ratio*

[61] *Krasner v Dennison (Lawrence v Lesser)* [2001] Ch 726.
[62] [2002] EWCA Civ 242, [2002] BPIR 847.
[63] Welfare Reform and Pensions Act 1999 s 11.
[64] [2000] 3 WLR 1117.
[65] Above.
[66] Trusts of Land and Appointment of Trustees Act 1996 ss 14 and 25.
[67] [2002] BPIR 556.

decidendi of the Court of Appeal[68] does not appear to impeach this particular interpretation.

The estate does not encompass the future income of the bankrupt.[69] The utilisation of this 'asset' is specifically reserved for an income payments order made under section 310 of the Insolvency Act 1986. It has to be said that this regime has been a statistical disappointment,[70] and when searching for reasons for that shortfall in expectations the role of the courts must be reviewed. It may well be the case that more income payments orders would have been granted had the pension rights of a bankrupt been viewed as future income. However, there is a consistent stream of authority running from *Re Landau*,[71] through *Krasner v Dennison*[72] to *Rowe v Sanders*,[73] to confirm that future pension rights must be seen as a current asset of the estate and thus automatically vest in the trustee without further ado. Apart from this limitation the courts have in modern times been generous in interpreting what is income (for example by including lump sum payments[74]) and in other ways in operating the section 310 regime.[75] At the end of the day the unpalatable truth may well be that only a small number of bankrupts will have a continuing income deemed worthy of being made subject to a section 310 application. That constituency has further narrowed with the advent of the IVA because such individuals have a reasonable prospect of persuading their creditors to go down the route of an arrangement rather than pressing for bankruptcy.

68 [2002] EWCA Civ 1174, [2002] 3 WLR 1760. The overturning of the first instance ruling was justified by the application of an old equity authority from the Court of Appeal (*Maclurcan v Maclurcan* (1897) 77 LT 474) suggesting that equitable rights could be created by unexecuted court orders. In the leading judgment of the Court of Appeal, Jonathan Parker LJ was placed in the invidious position of having to overturn one of his previous High Court judgments, namely *Beer v Higham* [1997] BPIR 349—see [2002] EWCA Civ 1174 at para 77 and [2002] 3 WLR 1760 at 1780—which he conceded had been handed down in ignorance of this binding authority. This conclusion has perturbed insolvency practitioners who fear that their actions in handling many similar cases will now be challenged.

69 *Supperstone v Lloyds Names Association Working Party* [1998] BPIR 832.

70 For a comprehensive review of s 310 in the courts see G Miller, 'Income Payments Orders' (2002) 18 IL & P 43. The statistics on income payments show that their usage has been disappointing—see Bankruptcy: A Fresh Start (Insolvency Service, London, 2000) at para 7.7.

71 [1998] Ch 223.

72 [2001] Ch 76.

73 [2002] EWCA Civ 242, [2002] BPIR 847.

74 *Kilvert v Flackett* [1998] BPIR 721. Compare this decision with the old (and now superseded) authority of *ex parte Benwell* (1884) 14 QBD 301 which limited the availability of income payment orders in cases where the bankrupt's income was fluctuating in nature.

75 See for example *Malcolm v Official Receiver* [1999] BPIR 97 (bankrupt told to leave expensive accommodation and to go downmarket to reduce mortgage commitments). Compare *Re Rayatt* [1998] BPIR 495 (court entitled to take into account cost of private education for bankrupt's children when assessing reasonable domestic needs of bankrupt).

VII. A MODERN APPROACH TO STEWARDSHIP

The idea that a bankrupt's affairs should be managed and realised by an independent professional has been at the heart of bankruptcy policy since 1883.[76] The introduction of compulsory professional qualification for insolvency practitioners was a further cultural change to the institution of stewardship, a precautionary measure only introduced in 1986. Having said that, one weakness in the legislation is the continued fiction that appointments are to be held by named individuals rather than vested in the firms from which those individuals are drawn. In practice it is clear that much of the day to day work is conducted by managers and not by the named partner. This has produced problems in commercial practice where it is increasingly common to find named partners in particular firms of insolvency practitioners moving to other firms whilst all of the files are retained by their previous firm. Similar problems arise on retirement where appointments have not been fully executed. The courts have been left to step into the breach to devise a pragmatic set of working rules governing the handing over of office.[77]

Clearly Parliament has placed various regulatory constraints upon the conduct of a trustee in bankruptcy as a steward. It is for example possible for a dissatisfied party (including the bankrupt) to complain to the court under section 303 of the Insolvency Act 1986. Such complaints have traditionally received short shrift[78] and the evidence even today is that they will more likely fail than succeed.[79] The courts simply do not like interfering with the exercise of management powers by a trustee.[80] In adopting this protective approach the courts have shown an understanding of the difficulties

[76] On the 1883 legislation see M Lester, *Victorian Insolvency* (Oxford, Clarendon Press, 1995) ch 5.

[77] See here first instance cases such as *Supperstone v Auger* [1999] BPIR 152, *Re Alt Landscapes Ltd* [1999] BPIR 459, *Re Equity Nominees* [2000] BCC 84, *Cork v Rolph, The Times* 21 December 2000, and *Clements v Udal* [2001] BPIR 454 where an evolving pattern of constructive and economy-conscious judicial decision-making in dealing with the practical aspects of resignation/retirement/replacement of insolvency practitioners is apparent.

[78] *Re A Debtor (No 400 of 1940)* [1949] Ch 236. This non interventionist approach dates back many centuries—see for example the protracted litigation involving the bankrupt King (1805) 11 Ves June 417, (1806) 13 Ves June 181 and (1808) 15 Ves June 127 as discussed by MS Servian, 'Coping with the Fraudulent Bankrupt: An Early 19th Century Perspective' (1987) 3 *Insolvency Law and Practice* 7.

[79] *Osborn v Cole* [1999] BPIR 251, *Hardy v Focus Insurance* [1999] BPIR 77. Compare *Re Cook* [1999] BPIR 881 and *Faryab v Smith* [2001] BPIR 246. In *Engel v Peri* [2002] EWHC 799 (Ch), [2002] BPIR 964 it was held by Ferris J that a bankrupt could invoke section 303 even after the date of his discharge. However, the use of s 303 to challenge the level of remuneration claimed by the trustee was thought to be inappropriate; a more suitable mechanism was to use the general power of the court under s 363(1) to control bankruptcy proceedings.

[80] On this see Harman J in *Port v Auger* [1994] 1 WLR 862 at 874 where the need to preserve the estate from the ravages of litigation costs was a reason cited to justify judicial non-interference. On the issue of interference with management note that the powers of the trustee with respect to the estate of the bankrupt survive the process of discharge. The unusual decision of

faced by insolvency practitioners and an awareness of the fact that by definition their actions will cause resentment in certain quarters. To allow such grievances to be litigated through the courts would inevitably lead to the estate being damaged further by unnecessary cost.

What Parliament has prescribed however is not the full story. The courts have in the past had, and continue today to have, an input here. Certainly, they have been at pains to assert the need for the trustee to retain his or her professional independence.[81] The fiduciary status of a trustee has also been stressed.[82] Most graphically, we could note the rule in *ex parte James*[83] that requires trustees in bankruptcy, by virtue of their status as officers of the court, to act 'honourably'. As James LJ put it:

> I am of the opinion that a trustee in bankruptcy is an officer of the court. He has inquisitorial powers given to him by the Court and the Court regards him as its officer and he is to hold money in his hands for its equitable distribution among the creditors. The Court, then finding that he has in his hands money which in equity belongs to someone else, ought to set example to the world by paying it to the person really entitled to it. In my opinion, the Court of Bankruptcy ought to be as honest as other people.[84]

The rule in *ex parte James*[85] is hardly ever applied in modern conditions, though the potential for its application remains.[86] More typically, the courts will decline any invitation to invoke it. Witness the case of *Mountney v Treharne*[87] where at first instance Stanley Burnton J indicated that for a trustee to seek to take advantage of his full legal rights for the benefit of the

DJ Caddick in *Wright v Official Receiver* [2001] BPIR 196 is best viewed as judicial unwillingness to interfere even at this stage. Here it was held that a debtor who had undergone the process could not escape the clutches of the trustee by seeking to promote an IVA at this stage. This county court case turned on the point that the former bankrupt was not a debtor after discharge and therefore could not qualify for an IVA.

[81] *Re Ng* [1997] BPIR 267 per Lightman J at 269—approved by the Court of Appeal in *Trustee of the Estate of Bukhari v Bukhari* [1999] BPIR 157.
[82] *Re Bulmer* [1937] Ch 499.
[83] (1874) 9 Ch App 609.
[84] *Ibid*, at 614. For a general review of the origins and applicability of this rule see I Dawson, 'The Administrator, Morality and The Court' [1996] JBL 437.
[85] Above.
[86] *Patel v Jones* [2001] EWCA Civ 779, [2001] BPIR 919. In *Upton v Taylor* [1999] BPIR 168 at 182 Rimer J indicated that the court could invoke this rule to control the remuneration of a trustee in bankruptcy. Note also the earlier cases of *Re Cornish* [1891] 1 QB 99, *Re Wigzell* [1921] 2 KB 835 and *Re Clarke* [1975] 1 All ER 453.
[87] [2002] BPIR 556 (Stanley Burnton J); [2002] EWCA Civ 1174, [2002] 3 WLR 1760 (Court of Appeal). In *Boorer v Trustee in Bankruptcy of Boorer* [2002] BPIR 21, Jacob J indicated that it was not unethical for a trustee in bankruptcy to take full advantage of his legal rights to protect the estate; in so doing he or she was entitled to change position in the event of new information (eg as to the value of property) coming to light even though such a change might disappoint a third party. See also *Re Gozzett* [1936] 1 All ER 79, *Re Byfield* [1982] 2 WLR 613, *Green v Satsangi* [1998] BPIR 55, *Walker v Hocking* [1998] BPIR 789, and *Re Ouvaroff* [1997] BPIR 712.

estate could not infringe that rule. The subsequent ruling[88] by the Court of Appeal that the trustee's rights were not as extensive as they may have appeared at first sight takes nothing away, it is submitted, from the general applicability of the comment of Stanley Burnton J. What is the reason for this desuetude of the rule? It may be because certain ethical standards are now embodied in legislation.[89] More probably, the advent of compulsory professional qualification, coupled with the requirements of professional codes of ethics, has rendered this rule of common law more marginal.[90] Advances in restitution law, particularly in the context of unjust enrichment, may also have resulted in the rule becoming otiose in a number of potential applications.

Of greater significance in modern conditions is the question of professional competence. There is no standard of competence explicitly set forth for trustees in bankruptcy. The Cork Committee called[91] for an explicit imposition of a duty of care but all we find in the statute is section 304 which purports to offer a remedy in the event of misfeasance or breach of duty by a trustee. No guidance is given in the legislation to amplify what is the required standard of competence. As far as trustees are concerned there has been limited jurisprudence and a marked reluctance on the part of the courts to find poor performance by trustees.[92] The most contentious authority concerns the position of the Official Receiver who in *Mond v Hyde*[93] was given an immunity in tort whilst discharging his public duties. This is a highly debateable judicial concession, particularly in the light of the requirements of the Human Rights Act 1998 which by incorporating Article 6 ECHR would arguably confer on an injured party a right to have his or her civil rights determined by a court without obstruction through a specially constructed doctrine of immunity.

It is important that trustees are equipped with appropriate powers to enable them to tackle their designated task efficiently. For the most part these powers are to be found in the governing legislation. Apart from the more obvious matters contained in Schedule 5 of the Insolvency Act 1986

[88] [2002] EWCA Civ 1174, [2002] 3 WLR 1760.

[89] See for example IR 1986 r 6.147 (controls on self dealing), r 6.148 (bar on touting for office).

[90] One should note here the policing role of JIMU (Joint Insolvency Monitoring Unit) and the more general remit of the Insolvency Practices Council—see D Harrison, 'The Insolvency Practices Council' [2002] *Ins Law* 175 for discussion of the latter body.

[91] Cmnd 8558 paras 777–8.

[92] See for confirmation of this fact *Green v Satsangi* [1998] BPIR 55 and *Brown v Beat* [2002] BPIR 421.

[93] [1999] QB 1097—for a critique see J Murphy, '*Mond v Hyde*: Negligence Immunity for the Official Receiver?' [1999] *Ins Law* 206. The case seems to have turned on the perception of an official receiver discharging functions as part of the administration of justice. In view of the increasingly deregulated nature of the bankruptcy procedure this assumption, based upon the earlier corporate insolvency authority in *Burr v Smith* [1909] 2 KB 306, is beginning to look out of touch with modern conditions.

the courts have sought to maximise available freedom of action. Thus, in *Stein v Blake*[94] the power of a trustee to assign a legal claim back to the bankrupt was upheld even though the consequence of so deciding was (because of the mandatory effect of bankruptcy set-off rules) to disadvantage a third party involved in litigation with the bankrupt. Lord Hoffmann conceded that the result of the Law Lords' ruling was unfortunate for the third party, but the problem lay in the rules governing legally aided litigation and should not be countered by the courts placing unnecessary restrictions upon the operational decisions of trustees in bankruptcy.[95] Generally, the courts are happy to see causes of action realised through assignment,[96] though the outmoded doctrine of champerty has the potential to present certain difficulties. Another noteworthy ruling on the extent of the trustee's powers is *Oakes v Simms*[97] where the Court of Appeal indicated that the power to undertake a private investigation pursuant to section 366 of the Insolvency Act 1986 survives the bankrupt's discharge.

VIII. THE IMPACT OF FUNDAMENTAL RIGHTS RECOGNITION

The issue of the fundamental rights in the bankruptcy arena is not new. But few would deny the added savour introduced into the menu by the Human Rights Act 1998. The courts have in the past two years had to tackle several key issues of potential conflict between established bankruptcy rules and the demands of fundamental rights legislation seeking to incorporate the European Convention on Human Rights and Fundamental Freedoms into English Law. Thus, the courts have been tested on issues of access to bankruptcy,[98] on the parameters of the estate,[99] on rights of spouses of bankrupts in respect of the family home[100] and on the lawfulness of arcane bankruptcy offences.[101] In *R v Kearns*[102] the Court of Appeal (Criminal Division) rejected an argument that the offence under section 354(3)(a) of

[94] [1996] AC 243. This case may be contrasted with *Faryab v Smith* [2001] BPIR 246 where the actions of a trustee in preferring to assign the cause of action to the defendant rather than back to the bankrupt was the subject of a successful challenge pursuant to s 303 of the 1986 Act.

[95] [1996] AC 243 at 261.

[96] The approach of the court in the unusual case of *Official Receiver v Davis* [1998] BPIR 771 is instructive.

[97] [1997] BPIR 499, following *Re Poulson* [1934] Ch 45.

[98] *R v Lord Chancellor ex parte Lightfoot* [2000] QB 597.

[99] See *Krasner v Dennison* (*Lawrence v Lesser)* [2001] Ch 76 and *Rowe v Sanders* [2002] EWCA Civ 242, [2002] BPIR 847 (Art 1 of Protocol 1 did not preclude the vesting of pension benefits in the estate).

[100] *Mountney v Treharne* [2002] BPIR 556 (reversed on different grounds—[2002] EWCA Civ 1174, [2002] 3 WLR 1760).

[101] *R v Muhamad*, [2002] EWCA Crim 1856, *The Times* 16 August 2002. But the courts are keen to limit the scope of this offence—see *R v P* [2000] 1 WLR 1568 (relevance of 'materiality' both as to cause of insolvency and quantum of losses emphasised by Court of Appeal).

[102] [2002] EWCA Crim 748, [2002] 1 WLR 2815.

the Insolvency Act 1986 of failing to account to the official receiver for loss of assets after the date of the bankruptcy petition infringed the right not to incriminate oneself. In delivering the judgment of the court Mr Justice Aikens indicated that this sanction was an appropriate penalty which was necessary to enable the estate of the bankrupt to be identified and protected. This conclusion is in line with other pronouncements on the question of the interface between bankruptcy law and fundamental rights. The usual outcome of such challenges is a judicial finding that the well established law is consistent with the European Convention on Human Rights and Fundamental Freedoms. This judicial determination has been reached on the basis that the rights conferred by the Convention are not absolute and that the Convention permits proportionate interference with such rights provided that is properly sanctioned by the law. These pronouncements have reassured insolvency practitioners who feared that the 1998 Act would unduly disrupt their day to day operations.

The English courts, then, have steered a cautious course. They are not unique in this conservative method of navigation. For in *Foxley v UK*[103] the European Court of Human Rights indicated that the power of an official receiver to seek a court order to redirect and open a bankrupt's mail pursuant to section 371 of the Insolvency Act 1986 did not infringe the Article 8 right of privacy. It was a proportionate tool, authorised by court order and designed to achieve a legitimate social purpose. However, interference with correspondence from the bankrupt's lawyers was not acceptable.

IX. THE NEW ALTERNATIVE TO BANKRUPTCY

The individual voluntary arrangement mechanism was new in 1985. It differed significantly from the old deeds of arrangement procedure (which, incidentally, was allowed to remain theoretically available) and so the jurisprudence developed under that mechanism was of limited utility. The deeds of arrangement regime was more heavily prescribed through primary and secondary legislation than was the case for the IVA at its inception; thus significant judicial insights on deeds of arrangement are necessarily circumscribed. The IVA was not modelled on any system in a comparable jurisdiction and therefore the courts could not draw support from the experience of their counterparts in other jurisdictions. They were on their own. How have they fared?

In terms of availability, the courts seem prepared to treat a proposal as falling within the voluntary arrangements scheme set forth in Part VIII of the Insolvency Act 1986 even though there may be questions as to whether

[103] [2000] BPIR 1009.

creditors will benefit from what is being proposed. Provided the required majority of creditors approve the proposal in good faith there is a disinclination on the part of the courts to double-guess the wisdom of their decision.[104] This is a practical conclusion for it is not the role of the courts to make commercial judgements on such matters; creditors are better placed to undertake such an exercise.

The legislative structure for the IVA has until recently been modest. But where the statute prescribes a requirement the courts have indicated that they are not inclined to exercise discretion to water down procedures for the simple reason that if the proposal is approved by the requisite majority dissenters will be bound. Thus, in *Mytre Investments v Reynolds (No 2)*[105] Blackburne J emphatically resisted the possibility of introducing a flexible concept of 'substantial compliance' to mitigate the precise rules on the giving of notice to creditors. To have decided otherwise would result in undesirable uncertainty being introduced into insolvency practice.

Clearly the legislation prescribes how an IVA may be initiated but it does not convey the full prerequisites. In *Cadbury Schweppes plc v Somji*[106] the Court of Appeal indicated that in addition to the legislative requirements there are overriding duties of good faith on the part of debtors proposing an arrangement. These duties require debtors to correct any information contained in their proposal which may have become misleading by the time at which the proposal is voted upon. This judgment will do much to instil confidence in creditors and will deter debtors and their advisors from promoting proposals that are presenting an incomplete picture of the financial position: such proposals are prime candidates for subsequent failure with all of the difficulties that such a consequence would produce for the parties and the insolvency system.

An early bone of contention concerned the issue of disputed voting entitlement. The exemplar here is *Doorbar v Alltime Securities*[107] where the pragmatism of the courts shone through. The problem here concerned

[104] See here the CVA case of *IRC v Adam and Partners Ltd* [2000] BPIR 986. Although there may be a *laissez faire* approach towards creditors who wish to take advantage of voluntary arrangements in an effort to recoup their debts, the courts do not have the power to extend the IVA procedure to debtors who do not qualify—see *Fletcher v Vooght* [2000] BPIR 435 where Lloyd J indicated that an IVA could not come into existence without a preliminary interim order.

[105] [1996] BPIR 464. The bankruptcy courts have recently cracked down on insolvency practitioners holding creditor meetings to vote on IVAs at a location not previously notified to the court—see *Re N (A Debtor)* [2002] BPIR 1024 (Mr Registrar Baister).

[106] [2001] 1 WLR 615. It is interesting how the first instance ruling in this case [2000] BPIR 950 was informed by old decisions on deeds of arrangement. On appeal the Court of Appeal indicated that old authorities should be treated with caution but that it would not be surprising if old principles were replicated in the IVA model. See also *Re Tack* [2000] BPIR 164 (Rimer J). This desire to protect the integrity of the interim order/IVA mechanism can also be seen in cases such as *Re Cove* [1990] 1 All ER 949, *Re Symes* [1996] BCC 137 and *Hurst v Bennett (No 2)* [2001] EWCA Civ 1398, [2002] BPIR 102.

[107] [1995] BCC 1149.

the question of unascertained debts and how they were to be treated by the chair of the creditors' meeting for voting purposes. The legislation requires the chair to *agree* a value for the unascertained claim. Does that refer to a bilateral process requiring the assent of the claimant or does it mean that the chair can conduct this assessment exercise unilaterally? The former conclusion would produce severe difficulties for practitioners and fortunately the Court of Appeal eschewed it.

In so deciding the Court of Appeal clearly was aware of the dilemma faced by the insolvency practitioner in dealing with claims that were speculative. Only wholly unreasonable valuations of claims would be liable to successful challenge.

Another area of difficulty left unresolved in the legislative scheme of things was the question of variation of IVA deeds. Parliament seems to have assumed that the arrangement would be agreed and then carried out to completion whether that be successful or unsuccessful. In spite of the convenience in being able to vary the terms of an IVA the courts, as typified by the approach of the Court of Appeal in *Raja v Rubin*,[108] have refused to allocate to themselves a remedial jurisdiction. The only possibility for variation outside the terms of the original arrangement is if all interested parties to the arrangement agree or the court makes and order under s 262 of the Act directing reconsideration of proposals. Thus the courts have asserted a contractual perspective in operating the IVA scheme.[109] Although in refusing to intervene the courts may appear to be unhelpful, in the long term this approach is beneficial in averting expensive applications to the court in cases where the matter should have been dealt with *ex ante* by prudent documentation.

Part VIII of the Act says nothing of standards of competence expected of nominees and supervisors. This may well be because of the attenuated nature of their stewardship (though we have noted a similar lacuna with regard to trustees in bankruptcy). Nevertheless, the courts have been required to fill this legal vacuum. So, for instance, in order to prevent weak proposals from coming before the courts a body of principle has been developed by the judiciary to the effect that nominees must be expected to undertake some form of viability assessment for the proposal.[110] Once the arrangement is in place what are the obligations of the supervisor? In *Pitt v Mond*[111] His Honour Judge Roger Cooke (sitting as a Deputy High Court judge) ruled

[108] [1999] 3 WLR 606. For the utility of a variation clause note *Horrocks v Broome* [1999] BPIR 66.

[109] This contractual perspective is apparent in *Strongmaster Ltd v Kaye* [2002] EWHC 444 (Ch), [2002] BPIR 1259 where Park J indicated that if an IVA had by its terms irretrievably expired that was the end of the matter.

[110] *Cooper v Fearnley* [1997] BPIR 20, *Greystoke v Hamilton-Smith* [1997] BPIR 24, *Hook v Jewson Ltd* [1997] BCC 752, *Knowles v Coutts & Co* [1998] BPIR 96, *Harmony Carpets v Chaffin-Laird* [2000] BPIR 61.

[111] [2001] BPIR 624. See also *Heritage Joinery v Krasner* [1999] BPIR 683.

that an insolvency practitioner was not to be held responsible for a wholly unpredictable collapse in property prices that rendered the mathematics of the proposal accepted by creditors as the basis for an IVA unviable. This case is particularly interesting for the perceptive analysis of the true nature of a voluntary arrangement and the precise role of a supervisor in such an arrangement:

> The juridicial basis of an IVA lies in contract, ie all creditors within the arrangement are bound by it .... The supervisor's function is in effect to supervise the carrying of the contract into effect and to petition for bankruptcy if the contract is not performed. What he is not as I understand the position is some form of receiver or manager.[112]

Commenting on the role of the licensed insolvency practitioner (LIP) as supervisor, Judge Cooke continued:

> It will readily be seen ... that the LIP 'changes hats' as matters proceed. Thus he may well start as the 'friend' and advisor of the insolvent but as matters progress he may lose that role completely and become someone whose concern is for the creditors and/or the public and therefore potentially 'the enemy'.[113]

More generally he observed:

> But it must be remembered that LIPs as a profession are required to be tough and unyielding people. Very often their task is to extract money from people who are unwilling to disgorge it; it means not taking 'no' for an answer. It is probably not a profession for the meek or fainthearted.[114]

Again in *King v Anthony*[115] an attempt by creditors to sue a supervisor for breach of statutory duty was scotched by the Court of Appeal; there was no such private right of action conferred by the law, merely the possibilities of seeking relief under section 263(3) of the Insolvency Act 1986 on the grounds of dissatisfaction, or of asking the court to exercise its inherent jurisdiction to regulate the actions of its officers. Criticisms of the terms of the arrangement or of the conduct of the meeting leading to its approval are best aired under section 262 of the Act. Having said that, it is clear that in cases involving section 262 petitions the courts do display an awareness for the inherent uncertainties faced by insolvency practitioners.[116] Such petitions rarely succeed; an outcome that is consistent with a policy of judicial

[112] [2001] BPIR 624 at 632.
[113] *Ibid*, at 633.
[114] *Ibid*, at 634.
[115] [1999] BPIR 73.
[116] For successful s 262 petitions see for example *Re A Debtor (No 222 of 1990) ex parte* Bank *of Ireland (No 2)* [1993] BCLC 233; *Re A Debtor (No 1 of 1999)* [2000] BPIR 998.

non-intervention in bankruptcy cases generally where the actions of a trustee are being questioned (for example, on section 303 applications).

The courts have been keen to facilitate supervisors who wish to exercise their statutory power to terminate an IVA.[117] They have recognised the importance of bringing failing arrangements to a speedy end and have buttressed the exercise of professional discretion in adopting that stance. A degree of institutional probity has thus been maintained and this is important for public and commercial perceptions of the voluntary arrangement process.

How have the courts addressed the problem of the unsuccessful IVA that peters out and is followed by bankruptcy? What happens to the IVA, and, more importantly, who gets the funds collected by the supervisor as part of that IVA? These questions have troubled the courts in a number of contradictory first instance decisions.[118] Some order was introduced by the Court of Appeal in *Re NT Gallagher & Son Ltd*,[119] a case dealing on its facts with a company voluntary arrangement but resolved by the application of rules also made explicitly applicable to IVAs. Under this important ruling the courts have stressed that the issue of termination is to be resolved by reference in the first instance to the provisions of the arrangement. Equally, the ownership of the IVA funds is to be governed by such matters. Most importantly, if the IVA fails to make appropriate provision the default rule is that bankruptcy does not bring the arrangement to an end and does not result in IVA funds passing from the supervisor to the trustee in bankruptcy for the benefit of creditors at large. As Peter Gibson LJ put it:

> Further, as a matter of policy, in the absence of any provision in the CVA as to what should happen to trust assets on the liquidation of the company, the court should prefer a default rule which furthers rather than hinders what might be taken to be the statutory purpose of Part I of the Act. Parliament plainly intended to encourage companies and creditors to enter into CVAs so as to provide creditors with a means of recovering what they

Note also *Ing Lease (UK) Ltd v Griswold* [1998] BCC 905 as a rare example of the court giving a direction to a supervisor under s 263(3) as to the operation of the IVA. Unsuccessful s 262 petitions were reported in cases such as *Re Naeem* [1990] 1 WLR 48, *National Westminster Bank v Scher* [1998] BPIR 224, *Re A Debtor (No 259 of 1990)* [1992] 1 WLR 226 and *Doorbar v Alltime Securities Ltd* [1995] BCC 1149.

[117] *Vadher v Weisgard* [1998] BPIR 295 (supervisor's decision to seek termination supported), *Re Keenan* [1998] BPIR 205 (proof of culpability by debtor not required), *Harris v Gross* [2001] BPIR 586 (supervisor could terminate even after scheduled date for expiry of IVA had passed), *Carter-Knight v Peat* [2000] BPIR 968 (default by debtor need not be continuing at date of hearing in order for court to act).

[118] *Re McKeen* [1995] BCC 412, *Davis v Martin-Sklan* [1995] BCC 1122, *Re Bradley-Hole* [1995] 1 WLR 1097, *Re Essengin Hussein* [1996] BPIR 160, *Kings v Cleghorn* [1998] BPIR 463, *Re Coath* [2000] BPIR 981.

[119] [2002] EWCA Civ 404, [2002] 1 WLR 2380. For the applicability of this case to IVAs see [2002] EWCA Civ 404 para 54 and [2002] 1 WLR 2380 at 2396.

> are owed without recourse to the more expensive means provided by winding up or administration, thereby giving many companies the opportunity to continue to trade.[120]

This watershed authority will engender confidence in those seeking to promote voluntary arrangements by ensuring that funds collected will be reserved for the exclusive benefit of VA participants. On the other hand, ironically, one consequence may be to make the operation of voluntary arrangements more difficult because new creditors will be placed in a position of greater jeopardy. They may not wish to take the risk of supporting a debtor in distress or will only do so at a prohibitive cost to the debtor. The problem of rescue funding requiring a super priority thus rears its ugly head. This is a major issue for policymakers (particularly in the corporate insolvency context) and with the best will in the world it is beyond the creativity of the judiciary.

X. APPRAISING THE JUDICIAL ROLE

One can conclude that the judiciary have been relatively supportive in promoting the goals prescribed by the Cork Committee[121] and enacted in the 1986 Act. Few cases can be cited of examples of what may be thought of as destructive judicial decision making. This broad consensus may well be linked to the cohesive nature of the judiciary operating in Chancery in bankruptcy cases.[122] Difficulties have however arisen in cases where those goals (and other dominant legislative/policy goals unrelated to bankruptcy) have come into conflict in particular circumstances. The legislation contains no hierarchical system of priorities and the courts have been left to choose an appropriate prioritisation. There is no consistent pattern of choice to be detected in the contest between competing bankruptcy goals. In some cases rules which aim to promote the creditor or public interest have won over rules seeking to protect debtors.[123] In other situations the debtor's expectations have prevailed.[124] A mixed picture also emerges in

[120] [2002] EWCA Civ 404 at para 50, [2002] 1 WLR 2380 at 2395.

[121] Cmnd 8558.

[122] Sir John Vinelott noted this cohesive quality in his essay, 'The Insolvency Acts 1985 and 1986' in [1987] *Current Legal Problems* 1 at 22–23 where the fact that the Cork recommendation for a specialist Insolvency Court was never implemented was commented upon.

[123] See for example *Cork v Rawlins* [2001] EWCA Civ 202, [2001] Ch 792. The case of *Re Saunders* [1997] Ch 60 could be noted here—claimants allowed retrospectively to commence proceedings against a bankrupt notwithstanding earlier failure to obtain leave of the court. In so deciding Lindsay J undertook a thorough review of the authorities (including Commonwealth decisions) and carefully weighed the policy of the law with regard to protecting assets of bankrupts against the needs of justice. The role of the court in controlling proceedings involving bankrupts was also very much to the fore.

[124] The best illustration is provided by *Haig v Aitken* [2000] 3 WLR 1117. The decision in *Solomons v Williams* [2001] BPIR 1123 (trustee failed to claim after-acquired assets within specified timeframe) may also be said to represent this limited genre.

the wider conflict between bankruptcy goals and extrinsic policy trends. There have been real difficulties in the interface with family law but the requirements of the Human Rights Act 1998 have not proved as testing as one might have imagined.

As far as the facilitation of commercial practice is concerned there is a good story to tell. Although no one would suggest that the courts are in the pockets of practitioners they have shown themselves acutely aware of the sensitivities of cases before them. One factor behind this empathy is the willingness of the courts to justify decisions by reference to the views of the Cork Committee, a body in which the practitioner constituency was very much to the fore.

The courts have also shown themselves adept in shaping bankruptcy law to changing social and technological conditions. For example, the now passé idea of voting by fax was quickly embraced by the courts.[125] In cases where they have baulked at change they have done so reluctantly and in a way designed to draw the attention of the legislature to the problem so as to enable corrective action.[126]

Moving from the general to the particular, the most constructive contribution of the courts has been with regard to the development of the IVA procedure. From a skeletal legislative structure the judiciary have constructed an impressive and sophisticated rehabilitation regime that has met with great success in practice.[127] In so doing they have displayed considerable acumen in grappling with the commercial and practical realities of the situation. With only tenuous statutory guidance they have constructed a foundation for the IVA which is firmly based in contract. In the long term and with intelligent IVA documentation this should reduce the flow of cases coming before the courts. They have emphasised that the supervisor's role is to 'police' that contract rather than to be involved in the day to day management of the debtor's affairs. They have firmly scotched the rumour that the supervisor is merely the poodle of the debtor; attempts by debtors to challenge what they perceive to be hostile actions by supervisors invariably fail. In so deciding the courts have limited their scope for future interventions.

A measure to evaluate the success of the courts can be found by a perusal of the reforms to personal insolvency law contained in the Insolvency Act 2000. There are provisions contained in this legislation which seek to neutralise

[125] See *IRC v Conbeer* [1996] BPIR 398, an IVA case heard by Laddie J.

[126] Reference may be made here to the case of *Re Austintel Ltd* [1997] 1 WLR 616 where the Court of Appeal upheld a first instance refusal of permission for bulk copying of insolvency records (particularly the files on company winding up petitions) for commercial purposes. In so doing they noted the complications created by technological advances in the keeping of court records and drew the attention of the Insolvency Rules Committee to the problem. Shortly thereafterwards bankruptcy files were opened up to the public—see SI 1999/359.

[127] The official insolvency statistics for England and Wales for the year 2000 show that there were 7978 IVAs as compared to 21,550 bankruptcies.

inconvenient case law[128] but for the most part the statutory reforms are intended to rectify errors in the 1986 legislative framework.[129] The changes in question in the 2000 Act deal with the new IVA procedure; bankruptcy with its more established heritage remains essentially untouched.

A similar test could be operated by looking at the personal insolvency provisions in the Enterprise Act 2002.[130] These reforms are not designed to counteract inconvenient judicial decisions; rather they represent a fundamental reappraisal of the social policy underpinning bankruptcy law. Thus, when this legislation takes effect we will have major relaxation in the law of bankruptcy with regard to discharge.[131] Automatic investigation of all bankrupts by the official receiver will no longer be the standard.[132] There will be a removal of disqualifications from honest bankrupts whilst their more disreputable colleagues will be liable to being subjected to bankruptcy restriction orders.[133] The power of trustees to realise the family home will be lost once a specified period has expired.[134] Income payments arrangements will be possible without the need for a court order.[135] Certain contentious bankruptcy offences will be abolished.[136] A significant expansion of the IVA regime is planned by allowing fast track arrangements for undischarged bankrupts; in such cases there will be no need for an interim order and the official receiver will be allowed to act as nominee/supervisor.[137] Unlike the Cork-inspired reforms these measures are controversial; they maintain the Cork tradition of liberalising the law in favour of bankrupts but are seen in many influential quarters as a step too far. This is particularly so when one recalls the concern of many commentators about the explosion of personal credit and then relates it to a suspicion that the government's preoccupation with promoting enterprise by alleviating the consequences of failure is missing the point about the causes of most modern bankruptcies

[128] It does for example neutralise the inconvenience caused by the necessary decisions in *McMullen v Cerrone* [1994] BCC 25 and in *Fletcher v Vooght* [2000] BPIR 435, but equally puts into statutory form standards of quality control developed in cases like *Cooper v Fearnley* [1997] BPIR 20. See s 3 and schedule 3.

[129] For example the 2000 Act addresses the difficulties caused by the drafting of s 252 in the areas of the availability and effect of an interim order.

[130] The Enterprise Act 2002 was introduced into Parliament via the Commons on 26 March 2002. It progressed to the Lords on 19 June 2002 and received the Royal Assent on 7 November 2002. The provisions dealing with personal insolvency law are not expected to be brought into effect before Spring 2004.

[131] The automatic discharge period will be reduced to one year—see s 256 which will replace s 279 of the Insolvency Act 1986.

[132] Enterprise Act 2002 s 258 (reconstituting s 289 of the 1986 Act).

[133] On BROs see s 257 (inserting s 281A into the 1986 Act).

[134] This was a late amendment to the Bill—see now s 261 (introducing s 283A into the 1986 Act).

[135] On this reform of the income payments system see Enterprise Act 2002 s 260 (introducing s 310A into the 1986 Act).

[136] Enterprise Act s 263—the 'retrospective' offences in the Insolvency Act 1986 of failing to keep proper accounts (s 361) and gambling (s 362) are abolished.

[137] See Enterprise Act 2002 s 264 and schedule 22.

which involve consumer debtors, who are not plunged into difficulty by the vicissitudes of enterprise, but rather by the temptations of readily available cheap credit.

The Enterprise Act 2002 will have some impact on the judicial role in bankruptcy. Central to the reforms contained in this new putative legislation will be the idea of a bankruptcy restriction order.[138] This concept is modelled on the company director disqualification order and will confer a potentially extensive jurisdiction on the courts in dealing with bankrupts who are deemed not to have been honest. Although there will be detailed statutory guidance on the exercise of this discretion one suspects that considerable judicial input will be required in the early years to establish a set of working rules. Certainly, that was the experience in director disqualification cases, which have provided the inspiration for this particular reform.

In the years to come the courts will be faced with a new challenge in adapting those statutory measures to fit the circumstances of individual cases. It will be interesting to see if the courts are as supportive of these reforms as they were of their predecessors some twenty years previously. This commentator suspects that the judicial contribution in the years to come will be more fractious. This is because the groundwork for the Enterprise Act 2002 has been less scrupulously researched than the reforms in 1985–6 and also because there is a greater range of commentators who have misgivings about some of the changes being proposed.[139] In researched than the reforms a climate of controversial bankruptcy legislation the performance of the courts in meeting the challenge of litigation will be critical.

[138] See schedule 20 to the Enterprise Act 2002, which inserts a new schedule 4A into the Insolvency Act 1986. Note that as part of a deregulatory strategy bankruptcy restrictions may be imposed by binding undertakings without the need for a court order.

[139] See Sir Gavin Lightman, 'Attitudes and Values: Developments in Insolvency Law (Part 1)' [2001] 14(8) Insolv Intell 57 at 60.

Part 4

Commercial Terms for Commercial Ends

14

Damages for Breach of Exclusive Jurisdiction Clauses

NIK YEO AND DANIEL TAN

I. INTRODUCTION[1]

AN AWARD OF damages for breach of an exclusive jurisdiction clause ('EJC')[2] has enormous, but as yet largely untested, potential. It could allow an aggrieved party to recover his costs of defending an action brought in a forum other than the chosen forum (the 'non-contractual forum'); possibly even allowing him to supplement any costs order made in his favour by an English court. It could also enable him to recoup any damages award made by a foreign, non-contractual forum in favour of the party breaching the EJC, effectively reversing that foreign court's decision; possibly also enabling him to top up any substantive award in his favour made by the non-contractual forum. These are radical possibilities. They raise a multitude of legal principles and issues of policy. Unless those principles and policies are applied with a weather eye on how they affect other situations in which damages for breach of an EJC may arise, the law risks developing inconsistently.

On normal domestic law principles of contract, a breach of an EJC should sound in damages. One might therefore have expected such actions to be commonplace. However that is not so. The existing case law, such as it is, will be discussed first. Two general objections to recovery for damages for breach of EJCs will then be touched on, before an attempt is made to map out the various factual circumstances in which damages may be sought, so as to provide some basis for the consistent development of this

[1] The comments of Adrian Briggs, Fellow and Tutor in Law, St Edmund Hall, Oxford, and Blackstone Chambers, Temple, and Yeo Tiong Min, Associate Professor, National University of Singapore, on an early draft of this chapter are gratefully acknowledged.

[2] For example, 'X and Y irrevocably agree that the courts of England shall have exclusive jurisdiction to hear and determine any suit, action or proceedings and to settle or hear any disputes which may arise out of or in connection with this Agreement'.

area of the law. In that mapping exercise some distinctions must be borne in mind. The principal one is between breach of an EJC in favour of an English court (a 'local EJC') and breach of an EJC in favour of a foreign forum (a 'foreign EJC').[3] Within those categories, two broad heads of damages exist: litigation costs (itself divisible into costs and expenses of pursuing pre-emptive remedies and costs of any substantive hearing) and recoupment of any substantive award made by the non-contractual forum. Various sub-categories are discussed later. The party in breach of an EJC will be referred to simply as 'X' and the party seeking damages as 'Y'.

It is assumed that the reader is familiar with the principles upon which English courts will stay their own proceedings brought in breach of a foreign EJC and when they will ordinarily grant an anti-suit injunction to enforce a local EJC.[4] In general terms, an English court will ordinarily grant a stay of proceedings commenced in breach of a foreign EJC *unless* 'strong reasons' or 'strong cause' can be shown why not.[5] In a similar fashion, an English court will grant an anti-suit injunction against X bringing proceedings in a foreign forum in breach of a local EJC unless 'strong reasons' or 'strong cause' is shown why not.[6]

This chapter deals with the common law position of damages for breach of EJCs and leaves for another time the effect that the Judgments Regulation[7] and the Brussels and Lugano Conventions have on the outcome, although, as English cases upholding the grant of anti-suit injunctions in relation to proceedings in other member states show, the common law position of remedies for breach of EJCs is likely to be at least highly relevant.[8] It is also assumed that English law governs the relevant EJC.

II. JUDICIAL VIEWS

A. EJCs

In the recent case of *Donohue v Armco Inc*,[9] Lord Bingham recorded, without any apparent disapproval, Lord Grabiner QC's concession that damages

[3] See O Kahn-Freund, 'Jurisdiction Agreements: Some Reflections' (1977) 26 *ICLQ* 825, at 827: it is necessary to maintain a clear distinction between local and foreign EJCs.
[4] For details see *Dicey & Morris on the Conflict of Laws* 13th edn (London, Sweet & Maxwell, 2000), Rule 32(2), at pp 424–31, 442–4, 448–50; Rule 31(5), at pp 386, 414–22.
[5] *The 'Eleftheria'* [1970] P 94, at 99–100. In that case, Brandon J set out various factors to take into account, including whether the applicant is merely seeking some procedural advantage in applying for a stay. However, as AS Bell argues, 'Jurisdiction and Arbitration Agreements in Transnational Contracts—Pt 1' (1996) 10 JCL 53, at 64–5, a procedural advantage foreseeable by the parties at the time should be excluded from consideration.
[6] *Donohue v Armco Inc* [2002] 1 Lloyd's Rep 425, at 433 & 439.
[7] Council Regulation (EC) No 44/2001 of 22 December 2000.
[8] *Turner v Grovit* [2002] 1 WLR 107 (subject to a reference to the ECJ); *Navigation Maritime Bulgare v Rustal Trading Ltd* [2002] 1 Lloyd's Rep 106, at 124.
[9] [2002] 1 Lloyd's Rep 425, at 437.

could be available for breach of a local EJC to recover any greater expense (for example, unrecovered litigation costs) or to recoup any greater liability in the foreign forum than would have been the case in England. Lord Hobhouse went further:[10]

> I am prepared to accept this submission and proceed on the basis that, if Mr Donohue can hereafter show that he has suffered loss as a result of the breach of the clause, the ordinary remedy in damages for breach of contract would be open to him. I say no more than this since the position is complex … . Lord Grabiner's point has merit and relevance in this exceptional and finely balanced case.

Lord Scott confined his comments to recovery of costs,[11] noting that he saw 'no reason in principle' why Donohue should not recover by way of damages for breach of the local EJC any costs of the New York proceedings incurred in a 'successful defence' of them. Lord Scott did not draw any distinction between costs of applying for any pre-emptive relief (a stay or dismissal, in the case of a foreign EJC; or an anti-suit injunction, in the case of a local EJC), and costs of fighting the substantive case brought against Donohue in the foreign forum.

Prior to *Donohue*, courts had already recognised that damages for breach of an EJC were, at least in theory, available.[12] However in *Continental Bank v Aeakos*,[13] Steyn LJ commented that damages for breach of a local EJC 'would be a relatively ineffective remedy'. That comment seems primarily to have been driven by the traditional pre-requisite for the granting of injunctions for breach of a positive covenant, that damages be an inadequate remedy.[14] It does not necessarily mean that, in principle, damages would not be available for breach of EJCs. The potential effectiveness of damages in this context has been underestimated.

Allowing a claim for damages for breach of EJCs may, of course, multiply litigation. In this context courts have noted that the risk of parallel litigation in two jurisdictions and the attendant risk of inconsistent decisions is

[10] [2002] 1 Lloyd's Rep 425, at 439.

[11] [2002] 1 Lloyd's Rep 425, at 443. Lord Scott's reference to 'costs' may have been to 'all liability and expenditure', rather than litigation costs in the narrow sense.

[12] *Continental Bank NA v Aeakos Compania Naviera SA* [1994] 1 WLR 588, at 598; *A/S D/S Svendborg v Wansa* [1996] 2 Lloyd's Rep 559, at 568. Although the latter decision was given on the 'strike out' standard—see Practice Direction [2001] 1 WLR 1001—Clarke J clearly did not think that damages were never available for breaches of EJCs.

[13] [1994] 1 WLR 588, at 598. See also *Angelic Grace* [1995] 1 Lloyd's Rep 87, at 96; *Schiffahrtsgesellschaft Detlev von Appen GmbH v Voest Alpine Intertrading GmbH* [1997] 1 Lloyd's Rep 179, at 190; [1997] 2 Lloyd's Rep 279, at 285; *OT Africa Line Ltd v Hijazy* [2001] 1 Lloyd's Rep 76, at 93.

[14] *London and Blackwall Railway Company v Cross* (1886) 31 Ch D 354, at 369; cf *Doherty v Allman* (1877–78) LR 3 App Cas 709, at 719–20.

an important, but not decisive, consideration.[15] Yet Lord Hobhouse in *Donohue* added the following qualification to his approval of Lord Grabiner QC's concession:

> when the issues of fact have been fully tried in New York, a situation may be established whereby Mr Donohue's right to rely upon the contract as against the defendants may be affected or situations of circuity of action may arise.[16]

This will be discussed later.

One case which recognises the existence of a right to damages for breach of an EJC is *Union Discount Co Ltd v Zoller*.[17] The Court of Appeal held that a claim for damages for breach of a local EJC, to recover the claimant's reasonable costs in successfully applying to have New York proceedings dismissed, should not have been struck out.[18] However the Court of Appeal stressed that its decision was confined to the facts before it, in particular that:[19]

(i) the costs were incurred in a foreign forum, defending proceedings brought against the claimant;
(ii) the foreign proceedings were in breach of a local EJC;
(iii) the foreign forum did not in these circumstances allow the successful party to recover costs; and
(iv) the foreign forum made no adjudication on costs.

The defendant conceded that breach of an EJC would in principle give rise to a liability in damages,[20] but argued against the recovery of foreign costs on various policy-related grounds. Of most relevance here, the court held that the English courts would not offend comity it they awarded, as damages, the costs of foreign litigation which exceed that recoverable in the foreign forum.[21] In so doing the court stressed the primacy of contractual rights: 'We see no policy reason *connected with either party* for allowing one party to the contract to escape from liability for the damages which he has caused to the other by attempting to sue [in breach of a local EJC] in a country where a different costs regime prevails.'[22] Concerns over parasitic litigation

[15] *Svendborg v Wansa* [1996] 2 Lloyd's Rep 559, at 570; *Credit Suisse First Boston (Europe) Ltd v MLC (Bermuda) Ltd* [1999] 1 Lloyd's Rep 767, at 781. Cf *Donohue v Armco* [2002] 1 Lloyd's Rep 425, 436.
[16] [2002] 1 Lloyd's Rep 425, at 439.
[17] [2002] 1 WLR 1517.
[18] Although the decision was on the 'strike out' standard, the Court of Appeal was clearly enunciating a new principle: see Practice Direction [2001] 1 WLR 1001.
[19] [2002] 1 WLR 1517, at 1524.
[20] [2002] 1 WLR 1517, at 1519 & 1524.
[21] [2002] 1 WLR 1517, at 1524–5.
[22] [2002] 1 WLR 1517, at 1526.

were also dismissed on the simple pragmatic ground that costs are often very much higher than the original sum in dispute.[23]

The Court of Appeal concluded by setting out what it regarded as 'more doubtful cases': where the foreign forum's costs regime was similar to that in England (eg in Australia); or where a party seeks to supplement by way of a damages claim the amount of a costs order made in its favour by an English court staying its proceedings because of an arbitration clause. These cases will be discussed later.

B. Arbitration Clauses

The question of damages for breach of arbitration clauses is relevant. Indeed, courts have acknowledged the similarity between clauses which require that disputes be referred to arbitration and not the courts ('arbitration clauses') and EJCs.[24]

The primary remedy today for breach of an arbitration clause is specific relief under the Arbitration Act 1996 to have proceedings stayed.[25] However, prior to statutory intervention, damages were the only remedy available for breach of an arbitration clause.[26] This perhaps explains why courts still readily acknowledge that damages are in principle also available for breaches of arbitration clauses, as did Browne LJ in *Mantovani v Carapelli SpA*.[27]

Yet courts have also recognised that a claim for damages as a result of X having obtained a foreign judgment in breach of an arbitration clause could require the chosen forum to determine whether it would have reached the same decision as the foreign forum—since Y would only be able to recover the amount by which the foreign award exceeds what would have been awarded by the chosen forum. In *Tracomin SA v Sudan Oil Seeds Co Ltd (No 1)*,[28] Sir John Donaldson MR described that as a 'rather unseemly spectacle'. This concern that the chosen forum has in effect to re-litigate the issue was echoed in *The 'Angelic Grace'*[29] and by Lord Hobhouse in *Donohue* (quoted above).

[23] *Ibid.*
[24] *The 'Lisboa'* [1980] 2 Lloyd's Rep 546, at 549; *The 'Angelic Grace'* [1995] 1 Lloyd's Rep 86, at 96 & 97.
[25] S 9 Arbitration Act 1996; *Doleman* [1912] 3 KB 257, at 267–8. And see *Heyman v Darwins Limited* [1942] AC 356, at 374; Mustill & Boyd, *The Law and Practice of Commercial Arbitration in England* 2nd edn (London, Butterworths, 1989) p 524.
[26] *Doleman & Sons v Ossett Corporation* [1912] 3 KB 257, at 267–8.
[27] [1980] 1 Lloyd's Rep 375, at 383; also at 384 *per* Megaw LJ; *Lisboa* [1980] 2 Lloyd's Rep 546 at 549 & 552; *The 'Atlantic Emperor'* [1992] 1 Lloyd's Rep 624, at 629, Neill LJ not disapproving of the trial judge's reference to damages for breach of an arbitration clause; *The 'Eastern Trader'* [1996] 2 Lloyd's Rep 585, at 600.
[28] [1983] 1 WLR 1026, at 1037.
[29] [1995] 1 Lloyd's Rep 87, at 94.

III. GENERAL OBJECTIONS

A. EJCs should Never Give Rise to a Right to Damages?

A solution to these difficulties is to say that an EJC is not a contractual term of the sort that gives rise to an action in damages:[30] an EJC does not give the parties rights *inter se*, but merely embodies a direction or request to courts. In this way, it is akin to a choice of law clause.[31] It is interesting to note that the Rome Convention specifically excludes EJCs from its scope,[32] which is at least consistent with the notion of EJCs being of a different *type* to other clauses in contracts; and arbitration clauses have been described as 'quite distinct from the other clauses' of a contract.[33] In *Mantovani*, counsel for the party who breached the arbitration clause argued that 'in the modern law the proper remedy for a breach, if there was one, was not to claim damages but to seek some form of equitable relief'.[34] That was rejected by the Court of Appeal and accordingly *Mantovani* stands for the proposition that damages are available for breach of an arbitration clause. In any case, this solution is far too broad, excluding the possibility of recovery for breach of EJCs in circumstances in which it has already been acknowledged (*Union Discount*) and in which it seems clearly justified.[35]

B. Is Quantification of Damages the Problem?

It has been suggested that the real problem with awarding damages for breach of an EJC lies with their quantification.[36] Issues of quantum are likely to cause most difficulty since X's liability for breach of an EJC is usually simple to establish: he has commenced proceedings in a non-contractual forum.[37]

However, English domestic law has never regarded difficulty of quantification as a reason to relieve the wrongdoer of having to pay

[30] Advocated by Look Chan Ho, 'Anti-Suit Injunctions in Cross-Border Insolvency: A Restatement' (2003) 52 *ICLQ* 697, at 707–710. See also A Briggs, Decisions of British Courts during 2001—Private International Law—note on *Union Discount* (2001) 72 *BYIL* 437, at 446.

[31] Which is arguably of this nature: A Dickinson, 'Restitution and incapacity: a choice of law solution?' [1997] *Rest LR* 66, at 69 fn 26.

[32] Art 1(2)(d) Rome Covention, in schedule 1 to the Contracts (Applicable Law) Act 1990.

[33] *Heyman v Darwins Ltd* [1942] AC 356, at 373; see also s 7 Arbitration Act 1996.

[34] [1980] 1 Lloyd's Rep 375, at 380.

[35] That is, at least recovery for foreign litigation costs where the foreign forum awarded Y some, but not all, of his costs (see s IV.A. below). However, as noted above, in *Union Discount* [2002] 1 WLR 1517, at 1519 & 1524, it was conceded that breach of an EJC did, in principle, sound in damages.

[36] WE Peel, 'Exclusive jurisdiction agreements: purity and pragmatism in the conflict of laws' [1998] *LMCLQ* 182, at 209 and 225.

[37] But see the discussion on what amounts to breach in s IV.F. below.

damages.[38] In any case, it is not quantification of damages *per se* which is problematic, but rather the consequences of quantifying them. Quantification is a matter for the *lex fori*[39] and English courts are well used to quantifying damages in cases involving the uncertainties of other potential litigation. For example, where a defendant seeks contribution from another towards an amount which the defendant has agreed to pay in settlement of litigation, courts have not been deterred by any difficulty in determining what would, in those circumstances, be a reasonable settlement.[40] In fact, where quantification of the precise loss is dependent on showing what a third party would have done, with all the attendant difficulties, courts have often resorted to quantifying the claimant's 'loss of chance'.[41] Where the claimant asserts that, as a result, say, of his lawyer's negligence, he has not been able to litigate his case properly, or at all, the court's quantification has taken the form of a loss of chance to litigate.[42] It is suggested below that this is a particularly powerful tool for quantification of damages for breach of EJCs.

***

Turning now to various factual situations, it will be seen that recovery, although potentially very broad, is not uncontroversial, notwithstanding that the two general objections just discussed are not persuasive. Existing principles of English domestic law bar recovery in certain of the situations. In others, there is undoubtedly cause for concern that allowing the recovery of damages is untoward, and it is suggested that other principles of English domestic law could be used to limit the right of recovery without having to rely on principles of private international law to cut down the scope of a fundamental domestic law right, the right to damages for breach of contract. There are, however, no neat, universal solutions to the difficulties.

The following discussion will be divided into breaches by X of local EJCs and breaches by X of foreign EJCs. Within each, certain fact situations will

[38] *Chaplin v Hicks* [1911] 2 KB 786, at 792.
[39] *Dicey & Morris on the Conflict of Laws,* at p 170.
[40] *Biggin v Permanite* [1951] 2 KB 314; *General Feeds Inc Panama v Slobodna Plovidba Yugoslavia* [1999] 1 Lloyd's Rep 688, albeit the burden of proving reasonableness is on the person seeking contribution.
[41] *Allied Maples Group Ltd v Simmons & Simmons (a firm)* [1995] 1 WLR 1602, at 1611; *Blue Circle Industries Ltd v Ministry of Defence* [1999] Ch 289, at 303–4. The claimant must establish as a matter of causation that he has a real or substantial chance of having suffered the claimed damages, not merely a speculative chance (*Allied Maples*). Nevertheless this is clearly a rough and ready approach. When a contestant is assessed as having a 20 per cent chance of winning, the award to her of 20 per cent of the £1,000 winner's prize is nothing if not pragmatic. She in fact had 0 per cent chance of winning £200 since the entire £1,000 was to go to the winner.
[42] *Kitchen v Royal Air Force Association* [1958] 1 WLR 563; *Peter Michael Harrison v Bloom Camillin (a firm)* [2000] Lloyd's Rep PN 89. Costs of that hypothetical litigation can be dealt with in the same way (*Harrison*). Whether a discount for the uncertainty in litigation will be applied depends on the level of uncertainty involved: *Charles v Hugh James Jones & Jenkins (a firm)* [2000] 1 WLR 1278.

be identified which illustrate the principal difficulties: where X breaches a local EJC by commencing proceedings in a foreign forum, Y may first seek to stay or dismiss the foreign proceedings (as in *Union Discount*) and Y may seek to recover losses he has incurred overseas. Further, he may seek an anti-suit injunction in the local forum, which he may or may not obtain—in each case potentially giving rise to costs in the local forum which he may seek to recover. Y may then contest the substantive action against him in the foreign forum and, if he is unsuccessful, he could suffer losses there.

Whether Y should be able to recover those losses by way of a damages action in England so as to unwind that foreign judgment is a most difficult issue and will be discussed at some length. Y may of course seek to counterclaim for such damages in the foreign proceedings, which will then be discussed briefly. Finally, the local court may refuse to accept jurisdiction over the substantive dispute, notwithstanding the local EJC. If the foreign forum then proceeds to hear the matter the difficult issue just alluded to arises again.

If X breaches a foreign EJC by commencing proceedings in the local forum, Y may obtain a stay and may incur expenses in the local forum not fully compensated for by an award of costs. If the stay is not granted and the substantive proceedings continue in the local forum, Y may suffer additional losses as a result of a substantive award made against him by the local court. That is a second controversial area. Finally, the three-forum situation will be discussed, where there is an EJC in favour of one foreign forum and X brings proceedings in a second foreign forum, with Y then seeking damages for breach in England.

IV. X BREACHES LOCAL EJC BY SUING IN A FOREIGN FORUM

A. Y does not Seek an Anti-Suit Injunction, But Successfully Challenges Jurisdiction of Foreign Court

Suppose Y is sued in a foreign forum in breach of a local EJC. He successfully challenges the jurisdiction of that foreign forum. On the authority of *Union Discount,* costs which Y incurs in the foreign court[43] can be recovered if the foreign court would not have awarded costs in favour of Y and so long as the foreign court had not even made any adjudication as to costs. Both factors should, however, be irrelevant.

As to the relevance of the foreign forum's costs regime: the Court of Appeal observed that Y may have agreed to the local EJC because of the costs regime in that local forum.[44] The point is broader. Y may have specifically

[43] Subject to normal principles of remoteness.
[44] [2002] 1 WLR 1517, at 1526.

wanted to avoid the procedural and costs regimes elsewhere, the parties having agreed to an *exclusive* jurisdiction clause. To give effect to this intention, as the Court of Appeal was willing to do, no distinction can be drawn between foreign forums on the basis of their costs regimes.[44a]

Suppose now that there was no blanket bar on Y recovering costs from the foreign forum. Should the fact that the foreign forum has made an adjudication on costs (either granting Y some costs or refusing costs altogether) debar Y from seeking to recover any shortfall in England in a damages claim? In *The 'Eastern Trader'*,[45] Rix J noted that he did not think any such claim had ever been successfully, if at all, advanced.

Although the Court of Appeal in *Union Discount* specifically confined its decision to where the foreign court had not adjudicated on costs, it held that there would be no problems of comity in an English court topping up a foreign costs award.[46] So the problem, if there is one, does not seem to be with comity. What about section 34 of the Civil Jurisdiction and Judgments Act 1982 ('CJJA')? It provides that no proceedings can be brought in England on 'a cause of action in respect of which' Y has obtained a judgment from a foreign forum against X unless that judgment is not enforceable or entitled to recognition in England.[47] The phrase 'same cause of action' is not dependent on the particular legal means by which an action is brought; rather, it will be satisfied if the factual basis of the causes of action are the same.[48] At a very general level, both a foreign costs award and Y's action for damages for breach share a common factual basis: Y has incurred costs in the foreign forum and that foreign forum has exercised its costs jurisdiction in favour of Y.[49] However that does not sufficiently explain what gives rise to Y's right to damages. The additional necessary fact forming the basis of Y's damages claim is that X sued in the foreign forum in breach of the local EJC. Hence it would seem arguable, and indeed desirable, that Y's action for damages does not constitute the 'same cause of action' as that 'in respect of which' that foreign costs order was made in his favour. So section 34 does not, therefore, apply to prevent Y's action for damages in England. If section 34 does not apply on its terms, it would be

[44a] See *A/S D/S Svendborg v Akar* [2003] EWHC 797, at para 37.
[45] [1996] 2 Lloyd's Rep 585, at 602.
[46] [2002] 1 WLR 1517, at 1524, although the Court of Appeal left open whether such an action would be stayed on public policy grounds (at 1527).
[47] The costs order will be enforceable against X in England since he will have submitted to the foreign jurisdiction by commencing proceedings there. S 32 and 33 CJJA will not apply.
[48] *Republic of India v India Steamship Co Ltd* [1993] AC 410, at 420.
[49] In absence of proof of how a foreign court would exercise its costs jurisdiction, English courts will apply English law as if it applied equally in the foreign forum: *Société Eram Shipping Co Ltd v Compagnie Internationale de Navigation* [2001] 2 Lloyd's Rep 627, at 637; cf *Dynamit AG v Rio Tinto Co Ltd* [1918] AC 260, at 295. Under English law, a costs order is made pursuant to a court's discretionary jurisdiction under s 51 of the Supreme Court Act 1981, which is exercised pursuant to the Civil Procedure Rules ('CPR'), particularly CPR 44.3.

odd for English public policy, stemming from a desire to prevent irreconcilable judgments, to bar Y from pursuing that action.

If Y would have been entitled to a foreign costs order but omitted to seek it, is X relieved of liability? In *Berry v British Transport Commission*,[50] Devlin LJ suggested that X would, on the basis that Y's loss was not solely a result of X's breach but a result of his own inaction. However, it is suggested that such an omission by Y could at most only assist X insofar as X could show that Y had failed to mitigate his loss by not seeking costs in the foreign forum. Use of causation is inappropriate in this context since it leads to an inflexible all-or-nothing result. If the foreign forum would not have awarded Y as much as he obtained in England, Y should still be able to recover the difference even if he did not seek such foreign costs. In any case, it would be unlikely that X could rely on Y's failure to mitigate for two reasons. First, any costs order, whether from a foreign or the local forum, would always be against X so there would be no practical reason why Y should be forced to litigate in the foreign forum.[51] Secondly, X (on whom the burden of proof lies)[52] may well find it hard to prove what the amount of the foreign costs order would have been, particularly if the foreign forum is a jurisdiction like England where the courts have a wide discretion on costs.

As a final matter, the Court of Appeal noted that Y's failure to seek an anti-suit injunction in the local forum does not bar his claim for damages; if relevant at all, it only goes to mitigation.[53]

B. Y Seeks and Obtains an Anti-Suit Injunction

Suppose Y seeks and obtains an anti-suit injunction in an English court. He may also obtain an order for his costs of that application, although probably only on the standard basis—that is, recovering only his reasonable costs which are proportionate to the matters in issue.[54] Y's actual costs may be in excess of that. Although an award of damages for breach of contract may, in extreme cases, be limited by notions of reasonableness and proportionality,[55] it is clear that any award of costs may well be less than what could fall

[50] [1962] 1 QB 306, at 321. The Court of Appeal in *Union Discount* noted that the fact the claimant did not ask for costs in New York was irrelevant since asking would have been fruitless: [2002] 1 WLR 1517, at 1527.
[51] If the foreign costs order would have been enforceable against X to a lesser extent than an English costs order (as a result, say, of X having fewer assets there) then Y should in any case be able to recover the difference.
[52] *Roper v Johnson* (1873) LR 8 CP 167, at 181–2 and 184.
[53] *Union Discount* [2002] 1 WLR 1517, at 1526.
[54] CPR 44.4.
[55] *Ruxley Electronics and Construction Ltd v Forsyth* [1996] AC 344.

within the bounds of recoverable damages, if Y were able to maintain such an action.

One immediately comes up against the principle in *The Quartz Hill Consolidated Gold Mining Co v Eyre*[56] that 'a party cannot claim by way of damages for those parts of the costs incurred [in a civil action] which he will not recover on taxation against his opponent either in the same action or in a separate [civil] action'.[57] In *Berry*,[58] Devlin LJ expressed regret[59] about this rule and said that, if the matter were *res integra*, he would allow Y to recover more than the costs awarded to him by an English court if he could show 'some additional ground for reimbursement over and above the bare fact that he has been successful'.[60] The matter was not, however, *res integra* and the Court of Appeal in *Berry* simply held that the rule did not preclude recovery in a civil action of unrecovered costs of a criminal action. In *Lonrho plc v Fayed (No 5)*,[61] *Berry* was taken as re-affirming that very rule and it was noted that the reason for Devlin LJ's regret of the rule (the fact that recoverable costs were at the time never likely to approximate a successful litigant's reasonable costs) had disappeared, since by the time of *Lonrho* a party awarded costs on the standard basis was entitled to all of his 'reasonable costs'. Since April 1999, the standard basis of costs has also been qualified by the requirement that the costs are 'proportional',[62] however that is not a reason to begin regretting the rule again, even less is it a reason to distinguish this line of authority.

In *Mantovani*[63] Megaw LJ recognised that where X commences proceedings in breach of an arbitration clause and Y obtains a stay of those proceedings 'there are no damages other than the costs of the application, which will be dealt with by the Court in the ordinary way'. That is clearly a reference to an award of costs pursuant to the court's discretionary powers. In not countenancing an action for damages for those costs, Megaw LJ implicitly supports application of *Quartz Hill* in these circumstances.

In the context of domestic law, the rule accords with justice and common sense. On the question of costs before an English court,[64] Y may make submissions relating to the merits of the case, including the fact that the

[56] (1883) 11 QBD 674.
[57] *Penn v Bristol & West Building Society* [1997] 1 WLR 1356, at 1364. See also *McGregor, McGregor on Damages* 16th edn (Sweet & Maxwell, London, 1997), p 501.
[58] [1962] 1 QB 306.
[59] See *per* Evans LJ in *Lonrho plc v Fayed (No 5)* [1993] 1 WLR 1489, at 1510.
[60] [1962] 1 QB 306, at 323. *Union Discount* [2002] 1 WLR 1517 took this as not precluding recovery of foreign costs which would not have been awarded to Y in any event.
[61] [1993] 1 WLR 1489, at 1510. See also *The 'Tiburon'* [1992] 2 Lloyd's Rep 26, at 33 (applying the *Quartz Hill* principle to a claim for costs against the first defendant which would otherwise have been recoverable as damages against the second defendant).
[62] CPR 44.4.
[63] [1980] 1 Lloyd's Rep 375, at 384.
[64] Submissions on how courts should exercise their discretion under s 51 of the Supreme Court Act 1981 and the CPR being made after judgment on the substantive issue.

proceedings were commenced in breach of a local EJC and, accordingly (but for *Quartz Hill*), Y could have expected to recover all of his reasonable costs by way of damages for breach. Y's submission is in effect for costs to be awarded to him on the 'indemnity basis' (that is, his reasonable costs, without any proportionality requirement, and with any doubt about reasonableness being resolved in Y's favour).[65] Y having made those submissions, and the court having made a determination on costs, it is inappropriate to allow that to be revisited in another action, even if Y has a free-standing right of action for breach of the EJC.[66]

By contrast, if Y did not raise X's breach of the EJC in any previous costs hearing and Y obtains costs on the standard basis only, is it then possible for Y to commence a fresh action for damages? There is some indirect support for such a proposition,[67] however, if this is right, Y would only be entitled to recover costs as assessed pursuant to the court's discretion on the indemnity basis.[68] Therefore, even though these may technically be fresh proceedings founded upon a distinct cause of action for damages, in practice, there would be a reopening of the costs hearing despite Y having previously had the opportunity to make submissions on X's breach of the EJC. It is suggested that Y should be disallowed from doing so on the basis of *Quartz Hill* or *Henderson v Henderson*.[69]

[65] CPR 44.3.

[66] Could Y avoid this bar by including in the EJC an express indemnity from X against loss suffered as a result of breach? It would seem not. CPR 48.3 provides that where a court is assessing costs in favour of a party who has a contractual right to them, then the contractual right is assumed to be limited to reasonable costs (unless it provides otherwise), and the court will assess them accordingly. Implicit in this is that quantification of any contractual right to costs will always be subject to the court's discretion (see *The 'Ikarian Reefer' (No 2)* [1999] 2 Lloyd's Rep 621, at 626). Although in *Gomba Holdings (UK) Ltd v Minories Finance Ltd (No 2)* [1993] Ch 171, at 192, the Court of Appeal commented that the discretionary power under s 51 Supreme Court Act does not affect any contractual right to costs, they allowed the appeal on the basis that the paying party always had the right to object to the reasonableness of costs claimed under that contract (in other words, that the contract did not provide a complete indemnity but at most allowed the payee to claim costs on the 'indemnity basis'). See also *Callery v Gray (Nos 1 and 2)* [2002] 1 WLR 2000, at 2025.

[67] This seems to be the case if Y had a contractual right to costs: *John v Price Waterhouse (a firm)* [2002] 1 WLR 953, at 961–2, although Ferris J noted that the point was 'novel' (at 964). Also *Gomba* [1993] Ch 171, at 194: (i) an order for costs is 'always a discretionary order' under s 51; (ii) that where there is a contractual right to costs, the discretion should ordinarily be exercised to reflect that contractual right; and (iii) that a mortgagee is not to be deprived of a contractual right to costs 'merely by reason of an order for payment of costs made without reference to the mortgagee's contractual or equitable rights and without any adjudication as to whether or not the mortgagee should be deprived of those costs'. See also Practice Direction as to Costs, 48 PD.1, at para 50.3(5).

[68] See the discussion in the last two footnotes, in particular point (iii) in the immediately preceding footnote which implies that a mortgagee could be deprived of his contractual right to costs if the court specifically addresses the issue (see also Practice Direction as to Costs, 48 PD.1, at para 50.3(4)(c)).

[69] (1843) 2 Hare 100: a party will be estopped from raising an issue in later litigation which could and should have been litigated in the earlier proceedings. That is just a species of abuse

Subject to that qualification, it seems that the general rule of English domestic law is that Y will be prevented from recovering damages for breach of an EJC in order to supplement (or potentially reverse) a costs order made by an English court (this rule, subject to that qualification, will be referred to as the '*Quartz Hill* principle').

C. Y Seeks and Fails to Obtain an Anti-Suit Injunction

The considerations in section IV.B. apply equally if Y fails to obtain an anti-suit injunction and costs are awarded against him. The *Quartz Hill* principle should still apply to bar Y from bringing an action for breach of the EJC in England so as to reverse that costs award. Y could have made submissions when it came to costs that, given X's breach of the EJC, X should in fact have to pay Y's costs. Costs could have been awarded to Y in such a case because English courts are not required to exercise their costs discretion in favour of the winning party.[70]

D. Foreign Court Nevertheless Proceeds to Determine Action, Awarding Damages Against Y

The mere fact that a foreign court has determined a substantive dispute instituted in breach (as an English court would have it) of an EJC will not preclude English courts from determining the same dispute so long as Y did not submit to that foreign forum, since the foreign judgment will not be entitled to recognition due to its being given in breach of that EJC.[71]

Recovery in an English action for damages of Y's costs of litigating in the foreign forum was specifically sanctioned by Lord Scott in *Donohue*.[72] If so, recovery is subject to the considerations discussed in section IV.A. above, subject to the following two qualifications. First, in this situation, Y may have submitted to that foreign forum to defend the substantive action. That could bar him from recovering these costs. Secondly, if (as will be suggested below as a possibility) Y should be restricted to recovering only his 'reliance loss', he will not be able to recover this category of litigation costs. These qualifications will be discussed in more detail later.

of process (*Johnson v Gore Wood & Co (a firm)* [2002] 2 AC 1, at 31), discussed in s IV.D.b. below.

[70] CPR 44.3(2)(b). See, for example, *Stocznia Gdanska SA v Latvian Shipping Company, The Times*, 25 May 2001.

[71] S 32 & 34 CJJA (the latter is only in any case relevant insofar as the foreign judgment was in Y's favour, and Y is seeking to recover more in the local forum). There is no issue estoppel here, since there is an independent issue in the English courts: breach of the EJC (cf *The 'Sennar' (No 2)* [1985] 1 Lloyd's Rep 521).

[72] [2002] 1 Lloyd's Rep 425, at 443. *Svendborg v Akar* [2003] EWHC 797, at para 37-8.

It is, however, recovery of any substantive award made against Y by the foreign non-contractual forum which is most controversial. Allowing Y in an English action for damages to recoup from X the very award made in X's favour by the foreign court reverses the effect of that judgment. It is arguably a more far-reaching remedy than an anti-suit injunction: whereas an anti-suit injunction is not enforceable overseas, a judgment against X in damages would, in general, be enforceable in all states bound by the Judgments Regulation[73] and the Brussels and Lugano Conventions. It therefore threatens property which X may have throughout Europe, and is likely to be especially more effective than an anti-suit injunction if England was chosen as a neutral forum (that is, a forum to which neither party has any particular connection).

Allied with an anti-suit injunction, which is essentially pre-emptive, an award of damages of this nature amounts to robust enforcement of EJCs. Because of its far-reaching consequences, it also has a prophylactic effect in discouraging Xs from attempting to breach EJCs in the future. A strong case can be made for allowing such damages. That case will be put forward first. Ultimately, however, broader public policy or comity concerns caution a more restrictive approach. It will be suggested below that there are options short of barring the right to damages altogether.

Although allowing such damages clearly has the effect of undoing the judgment of the foreign court, such an approach is not unknown to the law—see for example the 'clawback' provisions in section 6 of the Protection of Trading Interests Act 1980 which was enacted to protect UK companies from multiple damages awards from US courts given under anti-trust laws.[74] That legislation is clearly controversial, but what distinguishes the present proposal is that X will only be liable to such 'clawback' in our case because he breached an EJC which he expressly agreed to, whereas the Protection of Trading Interests Act extends to foreign awards within its scope even if the parties never contemplated that they would be affected by it. As the Court of Appeal noted in *Union Discount*,[75] the last person who should be allowed to benefit from any policy concerns which English courts have over reversing foreign judgments is the person, X, who has deliberately acted in breach of a local EJC—yet denying Y a right to damages in this case would do just that.

a. An 'Unseemly Spectacle'?

One particular concern should, however, be dealt with. It is the 'unseemly spectacle' mentioned by the Court of Appeal in *Tracomin* and *Angelic Grace*.

[73] Council Regulation (EC) No 44/2001 of 22 December 2000.
[74] See A Briggs, note on *Union Discount* (2001) 72 *BYIL* 437, at 446; L Collins 'Blocking and Clawback Statutes: The United Kingdom Approach—II' [1986] *JBL* 452, at 460ff.
[75] [2002] 1 WLR 1517, at 1522. See also *Berry* [1962] 1 QB 306, at 322–3.

What the Court of Appeal seems to have been referring to was the spectacle of the local forum deciding whether the foreign judgment was 'right or wrong.'[76] That, it is suggested, is a misrepresentation of the task. There may be all manner of rules in the foreign forum's private international law, or overriding statutes or public policy of that foreign forum, which compel the foreign court to a particular conclusion which the local court would not reach. The local forum's task is simply, as part of quantifying Y's loss, to determine how much Y would have been liable for if X had complied with the local EJC and sued here. The local forum is not required to speculate about the effect of foreign law, or what a foreign court might do. That is already known.[77] All that the local court has to decide is what *it* would have done. Lord Hobhouse's concern in *Donohue* with 'circuity of action' (insofar as it applies in this case)[78] can be addressed in the same way. The task of the local court is simply to determine the quantum of Y's loss in a separate action for damages. It will not create a vicious circle. It will be recalled that the Court of Appeal in *Union Discount*[79] held that the issue of whether Y should have applied for an anti-suit injunction went only to quantum. That implied a willingness of the courts, when determining quantum, to engage in the hypothetical exercise of asking whether they would have granted an anti-suit injunction had Y applied for one. A similar willingness is found in those (admittedly rare) cases where costs are determined even though the substantive dispute between the parties has become academic or has been resolved.[80] Such a task is nothing more than the sort of assessment courts carry out every day in determining the extent of a claimant's loss where he has been deprived of the ability to take or adequately defend legal action. As noted in section III.B. above, such issues of quantification have never traditionally distracted English courts.

It is a mistake to believe that the local courts may only satisfactorily fulfil their task by rehearing the substantive dispute in its entirety. The parties should be submitting evidence about what the local court would have done, rather than simply resubmitting the evidence which was before the foreign forum. The local forum can usually assume that it would have made the same findings of fact as the foreign forum and then simply assess what

[76] *Tracomin* [1983] 1 WLR 1026, at 1037.
[77] Pending exhaustion of all avenues of appeal in that foreign forum, assessment of damages should simply be postponed: see *Deeny v Gooda Walker Ltd (in liquidation)* [1995] 1 WLR 1206, at 1213ff. While Y could alternatively be granted an indemnity order by the English courts, the terms of that order would have to be crafted carefully to embody the relevant rules of remoteness of damage: see *Trans Trust SPRL v Danubian Trading Co Ltd* [1952] 2 QB 297, at 303 and 307. That makes this less attractive than postponing assessment.
[78] It applies with much greater force in cases of foreign EJCs, where a stay is not granted—see s V.B. below.
[79] [2002] 1 WLR 1517, at 1526.
[80] Eg *Boxall v Waltham Forest London Borough Council*, unreported, 21 December 2000 Scott Baker J (little doubt that the defendant would have been liable).

would have been the effects of applying the relevant legal principles to those facts. The only circumstances in which application of those principles is likely to yield a result which differs from that of the foreign court is (i) if the two forums apply different choice of law rules; (ii) if the two forums apply different overriding statutes or public policy doctrines; or (iii) if the two forums apply different procedural rules, for example, in relation to quantification or in relation to rules of evidence. In all those cases, except where rules of evidence differ, the local court should not need any further factual evidence. If, however, the evidential rules of the foreign forum were more restrictive than those in England (for example, a total prohibition on hearsay evidence in civil proceedings),[81] Y could argue that he would have been able to adduce further evidence in an English trial, and the court should assess what effect that would have had. In any case, even if further factual evidence were admitted, the English court has broad case management powers and could, for example, limit the length and scope of any cross-examination.[82] Unless the merits of the underlying case (through English eyes) are fairly clear, a preferable alternative approach would be to quantify Y's 'loss of chance' to have the matter resolved in England (see the discussion in section III.B. above). Although either approach necessarily duplicates the other to some extent, the question is whether court resources required to resolve such disputes are so great as to outweigh (i) Y's domestic law right to damages for breach of an EJC, and (ii) the desirability of upholding EJCs, which has recently been reaffirmed in *Angelic Grace*[83] and *Donohue*.[84] In this light, the 'spectacle' is perhaps not 'unseemly'.

b. Further Difficulties

Although the above concern can be addressed, unwinding a foreign award by way of a damages action in England is clearly controversial. It involves value judgments about the primacy of EJCs on which views may validly differ, and the reservations expressed in *Tracomin*, *Angelic Grace*, *Eastern Trader* and by Lord Hobhouse in *Donohue*, although not fully articulated, demonstrate the difficulty in simply assuming that the private law right of damages must always prevail. In addition to any feeling that such an action for damages is somehow untoward, the following two more specific difficulties should be noted. It will be observed that these difficulties are couched in general, policy-based terms.

First, it may well be thought that allowing such an action would bring the administration of justice into disrepute: it requires use of court

[81] Cf in England, s 1(1) Civil Evidence Act 1995.
[82] CPR 32.1(3).
[83] [1995] 1 Lloyd's Rep 87, at 96.
[84] [2002] 1 Lloyd's Rep 425, at 433 and 439.

resources in an admittedly complicated (albeit not insoluble) assessment of quantum simply to reverse a state of affairs which the English court may or may not have tried to enjoin by means of an anti-suit injunction. There is nothing necessarily unjust or irrational in insisting that, while Y may seek an anti-suit injunction, if it was not (for some 'strong reason') granted or if (as is more likely) it was not complied with and the foreign forum accepted jurisdiction, then Y would be barred from further compounding the problem of protracted satellite litigation by being able to reopen the dispute in England. An anti-suit injunction is primarily pre-emptive and does not attract the unnecessary duplication of resources which quantification of damages requires.[85] Y's action in pursuing such damages could therefore be regarded as an 'abuse of process'.[86] Although the categories of abuse of process are not closed,[87] it typically involves some sort of collateral attack on an existing judicial decision.[88] While formulated in the context of English domestic law, it has been applied in private international law, at least where there is fraud.[89] However, one wonders whether abuse of process itself is an appropriate tool for limiting Y's right to damages. Would it apply to bar unwinding the foreign substantive award but not recovery of any foreign litigation costs in defending that substantive action? What of foreign costs incurred in pre-emptive action? On what basis would such lines be drawn? Abuse of process is too subjective a basis on which to restrict such a fundamental domestic law right. It is only likely to satisfy those who already believe that the right to damages should be restricted in light of the broader concerns of private international law.

Secondly, an award of damages in this situation can create problems for those who believe in 'comity'. A useful way of determining whether problems of comity arise is to suppose the situation were reversed and ask what an English court's reaction would be.[90] Suppose Y and X had agreed to a foreign EJC which covered all relevant proceedings, but Y sought to sue X in England in the Employment Tribunal, relying on section 203 of the Employment Rights Act 1996 which renders EJCs void insofar as they purport to preclude recourse to Employment Tribunals. The Tribunal awards Y £1,000. Would the English courts not regard comity as being infringed if the

[85] '[T]o litigate about where you should have litigated ... seems even harder to congratulate than to litigate about where to litigate': A Briggs, note on *Union Discount* (2001) 72 *BYIL* 437, at 446.
[86] *Hunter v Chief Constable of the West Midlands Police* [1982] AC 529, at 536: it will be an abuse of process if Y's action would 'bring the administration of justice into disrepute among right-thinking people'.
[87] *Ibid.*
[88] *Reichel v Magrath* (1889) 14 App Cas 665, at 668. Cf *Hall v Simons* [2002] 1 AC 615, where abuse of process concerns were held not sufficient to warrant continuing the bar on negligence suits against barristers.
[89] *House of Spring Gardens Ltd v Waite* [1991] 1 QB 241, at 254–5. Cf *Turner v Grovit* [2002] 1 WLR 107, where abuse of process was used to *justify* a court's intervention.
[90] As the Court of Appeal did in *Union Discount* [2002] 1 WLR 1517, at 1524.

chosen foreign forum then allowed X to recover that £1,000 from Y on the basis of Y's breach of the EJC?[91] However, as with abuse of process, one wonders whether comity is itself an adequate tool to restrict the right to damages. Comity is clearly a somewhat vague and 'elastic' concept.[92] To rely on comity to limit the right to damages admits into contract law an unacceptable degree of uncertainty and, without providing sufficient independent justification, allows concepts drawn from private international law to 'trump' the domestic right to damages.

c. Limiting Techniques

If restriction of the right to damages is required by the issues which underlie the abuse of process and comity difficulties just discussed, and by the sentiments underlying the reservations in *Tracomin*, *Angelic Grace*, *Eastern Trader* and *Donohue* (*per* Lord Hobhouse), then it would be better to restrict that right by a technique more closely related to the domestic law of breach of contract. Any limiting technique drawn from English law clearly then depends for its effectiveness on the governing law of the EJC being English law.[93] However the problems of abuse of process and comity are not, it is suggested, as pronounced or therefore as objectionable if all the English court is doing is awarding damages for breach of an EJC pursuant to the domestic law of a foreign jurisdiction. (This will be discussed in relation to foreign EJCs in section V.B. below.) Possible techniques for limiting the right to damages will be explored in the remainder of this section IV.D.

d. Reliance Loss?

One contract-based technique for limiting Y's right to damages in this problematic situation would be to restrict his recovery to the 'reliance' measure of damages.[94] The aim of contractual damages is to put Y into the position in which he would have been in had the contract been duly performed,[95] and that is usually achieved by awarding Y the value of the benefits he would have obtained under that contract—the 'expectation measure'. However, Y seems to have a choice to seek to recover, in lieu of that measure

[91] German courts were apparently offended by interference with the effect of German consumer laws: *Phillip Alexander Securities and Futures Limited v Bamberger, Theele, Kefer, Riedel, Franz and Gilhaus* [1997] ILPr 73, at para 83.

[92] *Union Discount* [2002] 1 WLR 1517, at 1524.

[93] If the foreign governing law is not, as a matter of fact, proved before the English court then the court will apply English law as if it applied equally in the foreign forum: *Société Eram* [2001] 2 Lloyd's Rep 627, at 637.

[94] Thanks are due to Adrian Briggs for suggesting that this avenue be explored.

[95] *Robinson v Harman* (1848) 1 Ex 850, at 855.

of damages, losses incurred in reliance on the contract—the 'reliance measure'.[96] It has been said that the 'reliance measure' is tantamount to putting Y back into the position he would have been in had the relevant promise (here, the EJC) not been made—ie, the basis for compensation in tort,[97] although the similarity with the tortious basis may be purely coincidental.[98] To the extent that the reliance measure is equivalent to the tortious basis for compensation, the consequence of applying it in this situation is relatively clear, and the distinctions it draws are useful. The following losses would be likely to fall within the reliance measure: costs and expenses Y incurs in pursuing pre-emptive remedies (stays or anti-suit injunctions) to try to avoid having to contest the action brought by X in the non-contractual forum; and Y's costs and expenses incurred before X even begins proceedings (but at a time when the parties realised proceedings were likely) in preparing for what Y assumed would be an action brought against him in the chosen forum (such as costs of legal advice in that chosen forum about whether Y should make an early offer of settlement; costs of making evidence available in that chosen forum). These are all losses Y has incurred which he would not have incurred if the relevant promise (the EJC) had never been entered into. What would not be included are losses Y incurs as a result of the substantive proceedings brought against him in the non-contractual forum, which includes both recoupment of any substantive award made by that non-contractual forum, and recovery of costs of litigating the substantive hearing. They would fall outside the reliance measure because, even if the EJC had not been entered into, X would still presumably have brought proceedings in the non-contractual forum, it is just that X's action would not have been in breach of any EJC.

English cases on the reliance measure have been concerned with whether it is available to a claimant, rather than whether a claimant might be forced

[96] *CCC Films (London) Ltd v Impact Quadrant Films Ltd* [1985] QB 16, at 40; *Anglia Television Ltd v Reed* [1972] 1 QB 60, at 63–4. Cf *The Commonwealth of Australia v Amann Aviation Pty Ltd* (1992) 174 CLR 64, at 85–6, 105–7, 126–8, 136–7, 154–6, 162 (but cf at 166), where the High Court of Australia held that he did not have a choice, but that damages on this basis would only be available when the full expectation measure could not be valued in practice.

[97] See G Treitel, *The Law of Contract* 10th edn (London, Sweet & Maxwell, 1999), at p 875; McGregor, *McGregor on Damages* 16th ed (London, Sweet & Maxwell, 1997), at pp 30–1; *C&P Haulage v Middleton* [1983] 1 WLR 1461, at 1466; *Cullinane v British 'Rema' Manufacturing Co Ltd* [1954] 1 QB 292, at 303.

[98] *Amann* (1992) 174 CLR 64, at 82 and 85–6. See also at 107, 127–8, 136–7, 154, 162. Mason CJ and Dawson J noted (at 86) however that 'as a matter of strict logic' it is not possible to assess damages on that basis where expectation damages cannot be determined, and the majority of the High Court of Australia resolved this difficulty by assuming that any reliance expenditure would have been recouped had the contract been performed. This manifestation of the reliance measure does not assist in the present situation since it does not provide a clear reason for excluding recoupment of any substantive award made by a foreign court. In fact Deane J acknowledged (at 128) that there may be cases where reliance expenditure is incurred which would not otherwise have been incurred but for the breach; and that those cases might not be readily accommodated by the approach in *Amann*.

into adopting that basis and precluded from seeking his expectation measure.[99] A claimant will usually want to pursue the reliance measure when he has not proved or believes he cannot prove his expectation loss. Courts may be willing to compel him to adopt that measure if they conclude that there are 'serious' or 'insuperable difficulties' in determining his expectation loss,[100] however the circumstances in which the reliance measure can be imposed upon a claimant have never been considered in any detail by English courts.[101] In response to the concern over 'unseemliness' discussed above, it was argued that problems of quantification can be overcome. Yet, as noted, that depends on a particular view being taken of what resources are justifiably spent in the process. Y could be restricted to recovering his reliance loss on the basis that quantification of damages on the expectation measure is not practicable or economically efficient, or alternatively on the broader basis that to recover his full expectation loss is unreasonable and disproportionate,[102] creating, as it does, concerns of comity and abuse of process. This is a somewhat unusual application of the reliance measure, since it is normally resorted to where there is difficulty putting a 'cap' on Y's losses if he is allowed to claim the expectation measure. Here, however, the difficulty (such as there is) is in calculating the 'floor'—that is, the amount which Y would otherwise have been liable for had X sued him in the local forum. We already know what the foreign court has awarded (the 'cap'). The advantage with this technique is that it could be sensitive to the difference between recovery of costs in pre-emptive litigation and unwinding the effects of the substantive hearing in the foreign court. The difficulty is that English courts have yet to articulate a clear basis for imposing the reliance measure on claimants. The following technique, however, offers greater clarity.

e. Submission/Waiver?

In the fact situation being considered here, Y is likely to have submitted to the foreign forum and taken part in the substantive hearing. Clearly voluntary submission may debar Y from obtaining an anti-suit injunction, since it

[99] *Anglia Television Ltd v Reed* [1972] 1 QB 60; *CCC Films (London) Ltd v Impact Quadrant Films Ltd* [1985] QB 16.

[100] *McRae v Commonwealth Disposals Commission* (1951) 84 CLR 377, at 411–2 (High Court of Australia).

[101] Cf *Amann* (1992) 174 CLR 64 which held that reliance loss could only be resorted to when proof of expectation loss was practically impossible, however (as discussed in n 98 above) the court held that reliance loss was only a manifestation of the normal contractual basis for assessment, and this view of the reliance measure is therefore of no use in the present situation.

[102] See *Ruxley Electronics and Construction Ltd v Forsyth* [1996] AC 344.

may amount to a 'strong reason' to allow X to continue in breach of the local EJC.[103] In *Angelic Grace*,[104] Leggatt LJ implied that *Tracomin*-type damages would only be granted in the absence by Y of submission to the foreign forum; and in *Tracomin* itself there had been no submission. Moreover, Parliament's recognition of the importance of EJCs in section 32 of the CJJA (which provides that a judgment resulting from foreign proceedings brought in breach of an EJC is not enforceable in England) only applies if Y has not submitted to those foreign proceedings (section 32(1)(c)).

Rix J specifically addressed this issue in *Eastern Trader,*[105] and held that Y's submission to the foreign forum meant that he could no longer obtain damages for breach of an arbitration clause. Rix J commented that it was 'not easy to see why' a defendant should be able to 'relitigate' and claim damages when he could have obtained a stay under the Arbitration Act; and that allowing a damages action would, he thought, give the defendant 'a second bite of the cherry'. But the question of whether Y should have applied for a stay only goes to mitigation, as the Court of Appeal in *Union Discount* recognised in relation to anti-suit injunctions. Because a stay under section 9 of the Arbitration Act 1996 is virtually mandatory,[106] failure to apply for a stay in light of an arbitration clause is likely to be a good answer to Y's claim for damages. That justifies Rix J's conclusion, although it is not Y's submission *per se* that debars him, but his failure to mitigate. A failure to apply for an anti-suit injunction in light of a breach of an EJC will not so easily lead to the conclusion that Y has failed to mitigate. Moreover Rix J's comments about Y being able to 'relitigate' and have a 'second bite of the cherry' ignore the fact that Y should not have to (and should generally not be permitted to) present the same evidence as was before the foreign court; Y's task is simply to show that the local forum would have awarded a lesser amount, which the local forum can determine either precisely or on a 'loss of chance' approach. Y does not have a second chance to conduct his defence, but simply a (first) chance to restore his liability to the level it would have been had the EJC been complied with. The court is therefore simply compensating Y for his 'expectation loss'. However, is there any independent reason for baring a right to damages where Y has submitted to the foreign forum?

It is suggested that Y's submission to the foreign forum points to an acceptable basis for so doing, as it can be said to amount to a waiver by Y

[103] *Svendborg* [1996] 2 Lloyd's Rep 559, at 570.
[104] [1995] 1 Lloyd's Rep 87, at 94–5.
[105] [1996] 2 Lloyd's Rep 585, at 602.
[106] S 9 provides that a stay must be granted so long as Y can show that the dispute falls within the arbitration clause and that Y applied for a stay before taking any step in the proceedings to answer the substantive claim, unless X can show that the clause is void. However, as the question here is whether Y has taken reasonable steps to mitigate his loss, the burden of showing that Y could have obtained a stay would be on X.

of his rights to sue for damages arising after the date of that submission,[107] or possibly to an equitable estoppel preventing Y from so doing.[108] That would then preserve Y's right to sue for litigation costs incurred prior to his submission. There is nothing to prevent the parties agreeing to vary 'secondary obligations' to pay damages which arise upon breach of the 'primary obligations' under the EJC not to commence proceedings in a foreign court.[109] Any clear and unequivocal representation that Y does not intend to enforce some of X's secondary obligations should similarly be effective to give rise to a waiver or, possibly, an estoppel.[110] Y's submission could amount to such a representation. Such a limitation is clearly policy-driven but does reflect the concern that Y is having a 'second bite of the cherry'.

Yet it might be said that this technique is both too broad and, arguably, too narrow. It is too broad because it excludes recovery in the circumstances dealt with in section IV.A. where recovery of costs of taking pre-emptive action in the foreign forum seems clearly justified. At common law, any appearance by Y before the foreign forum to seek a stay or dismissal is likely itself to amount to submission for the purposes of recognition and enforcement of foreign judgments,[111] hence all losses incurred after that time would be irrecoverable. Section 33 CJJA reverses the position for that purpose. For submission to be an appropriate limiting technique, section 33 CJJA (or the equivalent principle) would need to apply in these circumstances. It is perhaps not difficult to conclude that, for the purposes of whether Y can be said to have waived his right to sue for damages, English courts would hold that only submission in the section 33 sense would be sufficient, since the highly technical notion of submission which prevailed at common law prior to that enactment is hardly something of which Y is likely to have been aware.

[107] In *DVA v Voest* [1997] 1 Lloyd's Rep 179, at 189, Morison J suggested that the reason for an anti-suit injunction not being available once Y has submitted is that 'it is tantamount to a waiver of the contractual provision on the basis of which the anti-suit injunction is founded' (appeal dismissed at [1997] 2 Lloyds Rep 279). See generally, *The 'Kanchenjunga'* [1990] 1 Lloyd's Rep 391, at 397–8; alternatively, *Banning v Right (Inspector of Taxes)* [1972] 1 WLR 972, at 979, which S Wilken & T Villiers, *Waiver, Variation and Estoppel* (John Wiley & Sons, Chichester, 1998), p 60, classify as 'pure waiver', for which a representation may not be needed. So far as this is waiver by election (*Kanchenjunga* waiver) it must be doubted whether Y would have knowledge of the right to choose between suing for damages and submitting, as required by *Peyman v Lanjani* [1985] 1 Ch 457. Perhaps Y's knowledge could be imputed (cf Wilken & Villiers, *op cit,* pp 57–9).

[108] See generally, *Thomas Hughes v The Directors etc of the Metropolitan Railway Company* (1877) 2 App Cas 439. The necessary element of X's reliance to his detriment could be constituted by X's continuing with the foreign proceedings.

[109] *Photo Production Ltd v Securicor Transport Ltd* [1980] AC 827, at 849 (although the secondary obligations cannot be excluded entirely). The EJC should, for these purposes, be regarded as a separate agreement, so that secondary obligations under it can arise independently of whether secondary obligations have arisen under the rest of the contract in which the EJC appears.

[110] See for example, *The 'Chemical Venture'* [1993] 1 Lloyd's Rep 508, at 521.

[111] *Henry v Geoprosco International Ltd* [1976] QB 726.

On the other hand, submission might be said to be too narrow since it does not bar recovery of damages by Y if Y has not submitted but merely suffered a default judgment in the foreign forum. However, in this case, concerns with Y recovering damages and unwinding a substantive foreign award are not as pronounced or as objectionable. After all, if X obtained a judgment in default of Y's appearance there is unlikely to have been any full trial in the foreign forum. Moreover, section 32 CJJA would, in these circumstances, apply to render that judgment unenforceable in England. Those very facts lessen the concerns over comity and abuse of process. Hence this limiting technique cannot be said to be too narrow.[112]

Y may nevertheless suffer substantial damages as a result of a foreign default judgment being enforced against his assets overseas. In an action for those damages brought in England by Y, could X plead that Y's omission to submit and defend the foreign proceedings amounted to a failure to take reasonable steps to mitigate Y's loss? Such a conclusion would be unusual,[113] although not totally inconsistent with the rules of mitigation.[114] Y's appearance and defence could be expected only to lessen the chance of an adverse finding being made against him and so, on one view, it could be a reasonable step for Y to take to mitigate his loss.[115] This would, however, require the English court to predict, not only what *it* would have done if X sued here, but what result would have been reached in the foreign forum had Y defended those proceedings. Although, at base, this is merely an evidentiary matter,[116] it does raise, in a heightened fashion, the abuse of process and comity concerns discussed above. It is suggested that X should not be able to rely on mitigation in these circumstances. One might say that, although sometimes a claimant can be required by the duty to mitigate to pursue litigation, litigating in a foreign non-contractual forum is simply too speculative;[117] alternatively, that it is simply unreasonable to require Y to

[112] If Y only had assets in England he would be unlikely to have any quantifiable loss. However that is a matter of chance and would by no means exclude Y from recovering in every case.

[113] Generally, it will be difficult for a defendant to show that a claimant did not take reasonable steps to mitigate since it is the defendant who is the wrongdoer (*Banco de Portugal v Waterlow and Sons Ltd* [1932] AC 452, at 506).

[114] Y may in certain circumstances be required to commence litigation against a third party in order to discharge his duty to mitigate (*Halifax plc v Gould & Swayne* [1999] PNLR 184, at 200–1). Defending an action brought in a foreign forum by X would generally be less onerous and expose Y to less uncertainty than an action against a third party, and so it should be no more difficult to conclude that Y has to submit and defend those proceedings in order to mitigate his loss.

[115] The fact that both foreign and local awards would be against the same person, X, would not itself relieve Y of the need to take such mitigating action since judgments of the foreign court are unlikely to be enforceable against Y to precisely the same extent as those of the local forum.

[116] Briggs & Rees, *Civil Jurisdiction and Judgments* (London, Informa, 2001), p 294, n 270.

[117] A claimant will generally not be required to take speculative legal action against third parties (*Pilkington v Wood* [1953] Ch 770).

do specifically what X and Y both contracted neither would do—ventilate their dispute in substantive proceedings in a foreign forum.[118]

In summary, therefore, it is suggested that the objections to recovering substantial damages manifested by the concerns over comity and abuse of process discussed above are adequately addressed by barring Y from recovering damages in circumstances where he has waived his domestic law right to damages (or is estopped in equity from asserting it). Whether Y took part in the substantive foreign proceedings would determine his ability to recoup any substantive award made against him (or top up any award made in his favour). It will also affect whether he may recover the costs of litigating those substantive proceedings.[118a] However, appearing before the foreign forum merely to contest jurisdiction should not amount to a waiver or estoppel, and the costs of doing so should, accordingly, be recoverable in an action for damages.

E. Y Counterclaims in the Foreign Forum for Breach of the EJC

If, prior to bringing proceedings in England for breach of the local EJC, Y counterclaimed in the foreign forum and obtained an award for damages for such breach, then it is clear his action here is barred by sections 32 and 34 of the CJJA.

F. The Local (English) Court Stays its Proceedings Notwithstanding a Local EJC

The English court may stay proceedings notwithstanding a local EJC that purports to confer on it exclusive jurisdiction. This can occur if the court concludes there are 'strong reasons' to allow litigation to proceed in a foreign forum, such as where the English action is related to other proceedings in that foreign forum.[119] In such a circumstance, would Y be barred from bringing an action in damages for breach of the EJC in England for the shortfall between what Y would have recovered in England and what he did recover in the foreign forum?

As a practical matter, one must assume here that Y commenced the English action and X applied for a stay. There must be some doubt as to

[118] The mere fact that Y is forced to deal further with X is not itself a reason for relieving him of the duty to mitigate (see, for example, *Payzu Ltd v Saunders* [1919] 2 KB 581). However here Y would be forced to deal with X in the very way which constitutes X's breach of contract.

[118a] Cf *Svendborg v Akar* [2003] EWHC 797, at para 37–8.

[119] For example, *Donohue* [2002] 1 Lloyd's Rep 425; *Credit Suisse First Boston (Europe) Ltd v MLC (Bermuda) Ltd (formerly MLC Emerging Markets Ltd)* [1999] 1 Lloyd's Rep 767; *Bouygues Offshore SA v Caspian Shipping Co (No 2)* [1997] 2 Lloyd's Rep 485 and [1998] 2 Lloyd's Rep 461.

whether X's action will always be a breach of the EJC. If the EJC merely requires that 'the courts of England shall have exclusive jurisdiction', does that embody an undertaking by X that he will not seek a stay of those local proceedings? After all, a stay at common law does not divest English courts of their jurisdiction.[120] It would seem sensible to construe any simple EJC[121] as not preventing X from seeking a stay of proceedings in the chosen forum, otherwise defendants could not even seek to challenge service of English proceedings out of the jurisdiction[122] where there was a local EJC without being in breach of that EJC. This position clearly cannot be correct. Where, however, the EJC also embodies an express promise not to challenge the jurisdiction of the chosen court, matters are different. Even if a breach of the EJC by X is demonstrated and X obtains a stay, Y would still have to demonstrate that his loss was caused by X's action in applying for the stay and not by the English court in granting the stay.[123]

Would the English court then also stay Y's action for damages? The fact that, say, there is parallel litigation in the foreign forum may well incline the court to stay Y's damages action, although that would not necessarily be so since the 'strong reasons' invoked by the court in originally staying the substantive action are essentially pragmatic reasons which may or may not also apply to Y's action for damages.

If the court accepts jurisdiction over Y's action, the same considerations set out in sections IV.B. and IV.D. apply. For example, if Y were to submit to the foreign forum and contest the substantive proceedings brought by X, he should not be able to recover costs of doing so or to recoup any substantive award, and his costs of resisting the stay in England would, it seems, be subject to the *Quartz Hill* principle referred to above.

V. X BREACHES A FOREIGN EJC BY SUING IN ENGLAND

A. Stay Granted in Local (English) Court

If X commenced proceedings against Y in England in breach of a foreign EJC and Y successfully applied for a stay and his costs, Y would not, it seems, be able to sue for breach of the EJC in order to recover any shortfall in his costs because of the *Quartz Hill* principle.

[120] *Empson v Smith* [1966] 1 QB 426.
[121] See n 2 above.
[122] That can only occur in the limited circumstances set out in CPR 6.19 and 6.20.
[123] Courts take a 'common sense' approach to causation in contract: *Galoo Ltd (in liquidation) v Bright Grahame Murray (a firm)* [1994] 1 WLR 1360. Since X's obligation is not to litigate in the foreign forum, it is likely that any award made by the foreign forum would be held to have been a natural consequence of, and hence caused by, X's breach, just as a lawyer's bad advice would be the cause of losses suffered by a company acting on that advice, notwithstanding the directors had a discretion whether or not to act on it: *British Racing Drivers' Club Ltd v Hextall Erskine & Co (a firm)* [1997] 1 BCLC 182.

B. Stay not Granted and Y Counterclaims for Breach of EJC in English Proceedings

As a result of the *Quartz Hill* principle, Y's costs[124] (and any costs award made against him), whether in relation to the stay hearing or in the substantive hearing in England, cannot be the subject of a subsequent action for breach of the EJC in England. It would seem that the *Quartz Hill* principle would similarly prevent Y counterclaiming in the relevant proceedings for breach of the EJC in relation to those costs. Although any costs award would not have been determined at the time of the counterclaim, Y could in theory obtain an order that X indemnify him in respect of the amount of those excess costs;[125] yet any such indemnity would only serve as guidance to a court when considering how to exercise its discretion as to costs, and would, in practice, preclude a further action on the indemnity.[126] Accordingly, even if the costs of an English trial would be greater than the costs of a trial in the chosen foreign forum, all Y can do is make those submissions before the English court when the time comes for the court to exercise its discretion as to costs. Those submissions can clearly be quite forceful since X has obtained an extraordinary acceptance of jurisdiction by the English court notwithstanding the foreign EJC. In those circumstances, Y could say it was unfair for him to be liable in England for costs over and above those for which he would have been liable in the chosen foreign jurisdiction.

The difficulty with this argument (even when made as part of a submission as to costs) is that it requires the English court to predict the decision a foreign court would have reached on the question of costs, had the substantive hearing occurred there, and (to the extent that the foreign forum's rules on costs depend on the outcome of the substantive matter) to predict the outcome of the substantive matter itself. This is a degree more difficult than the court's task in assessing what *it* would have done (which is the case where damages are sought to recover a foreign award made in breach of a local EJC—see section IV.D. above).[127]

What if the English court then makes a substantive award against Y or in favour of Y but in circumstances where Y believes that he would have

[124] If Y were restricted to the reliance measure, there would be certain of Y's reliance expenditure which would not be recoverable by way of a costs award, such as costs thrown away in preparing for a hearing in the chosen foreign forum. Recovery of those expenses would not seem to be barred by *Quartz Hill*—see, in the context of a contractual right to costs and expenses, *Gomba* [1993] Ch 171, at 192, where the Court of Appeal held that the s 51 discretion does not extend to expenses which could not form the subject of a costs award (what it called 'non-litigation costs').

[125] *Trans Trust SPRL v Danubian Trading Co Ltd* [1952] 2 QB 297, at 303 and 307.

[126] See s IV.B. above, in particular the footnote discussions of *Gomba*.

[127] Although at base it is just an evidentiary matter: Briggs & Rees, *Civil Jurisdiction and Judgments* (London, Informa, 2001), p 294, n 270.

fared better in the chosen foreign forum? The answer is generally that Y should not be able to recover, for two interdependent reasons.

First it would be inconsistent for the English court to entertain any action for breach of the EJC since that would have the practical effect of unwinding the award which it had just made. This is perhaps more what Lord Hobhouse refers to by undesirable 'circuity of action'.[128] Briggs and Rees recognise this argument but respond that where the English court declines to stay proceedings this does not preclude a right to damages, just as where, in domestic law, a court of equity's refusal to grant specific relief does not affect the claimant's common law right to sue for damages.[129] Although that is certainly correct as a general principle, it fails to take account of the peculiar private international law circumstances in which this issue arises—in particular, that the English courts will only refuse such a stay if 'strong reasons' are shown for doing so. To allow an action for damages would generally undermine the 'strong reasons' which the court had just relied upon to accept jurisdiction notwithstanding the foreign EJC. For example, suppose the English court were to decide to accept jurisdiction, notwithstanding a foreign EJC, because it considered X would not be able to obtain a fair trial in the chosen foreign forum; it awards X £10,000. To allow Y's counterclaim that, had the trial taken place in that foreign forum, X would only have been awarded £2,000 because the foreign forum discriminates against Xs would be self-defeating and, indeed, it would detract from the judicial discretion to hear a case, in appropriate circumstances, despite a foreign EJC. The point is not that there is necessarily jurisdictional inconsistency between refusing specific relief and still allowing a damages action, but simply that there is a practical inconsistency. On a more pragmatic level yet, it gives the same net result as if a stay had been granted in the first place, but at many times the litigation cost.

Secondly, however, certain types of 'strong reasons' may not in fact be undermined. If the ground for refusing a stay was that there were a multiplicity of related proceedings already taking place in England and resolution of X's claim against Y which is covered by the foreign EJC would be far more efficient if it were also carried out by the English courts,[130] then it is not inconsistent for the English court to conclude that, as between Y and X, their rights should be determined as if the matter had been heard by the chosen foreign forum, and to allow Y to counterclaim to reach that result.

[128] *Donohue* [2002] 1 Lloyd's Rep 425, at 439.

[129] Briggs & Rees, *Civil Jurisdiction and Judgments*, p 294. The second obstacle mentioned by Briggs & Rees, that the EJC is an illegal ouster of jurisdiction and hence cannot found an action for damages, is rightly dismissed by them. It should be noted that the authors acknowledge that whether Y has a right to damages 'must remain uncertain' (*ibid*, at p 295).

[130] Peel argues that in fact this is one of the only justifiable bases for a finding of 'strong reasons': WE Peel, 'Exclusive jurisdiction agreements: purity and pragmatism in the conflict of laws' [1998] LMCLQ 182, at 222.

The distinguishing factor here is that the 'strong reasons' involve other related litigation. Nevertheless, the concerns with unwinding judgments of the non-contractual forum discussed at section IV.D. might apply with even more force here, since the English court's task will be to determine, not what *it* would have done (the English court has already reached its decision), but what the chosen foreign forum would have done. Although that is not an impossible task, it does make the concerns expressed about apparent abuse of process and unjustifiable waste of court resources somewhat more pronounced. There is an added complication in seeking to debar Y from recovering damages in this fact situation: a foreign EJC may often not be governed by English law, and in those circumstances the limiting techniques discussed in section IV.D. will not be applicable—the English court will be required to apply the applicable *lex causa*. Where English law does govern, and where Y has submitted to the principal English proceedings brought by X, that should amount to a waiver of Y's right to sue for damages. In any case, where there is the practical inconsistency described above, the English courts should stay any counterclaim or subsequent action brought by Y.[131] In all other cases (such as where the English court has accepted jurisdiction over Y because of other related proceedings in England, and where the EJC is governed by a foreign law which would not otherwise debar Y or where it is governed by English law but Y did not submit to the principal English proceedings and only commenced a damages action subsequently), there would seem to be no reason of principle or practice why Y should be debarred from recovering any substantive award.

Recoupment of litigation costs in England would seem to be barred by the *Quartz Hill* principle. Y would be restricted to recovering any other losses and expense he may have incurred which are not properly costs of the English proceedings (such as preparing for a trial which he had been expecting would occur in the chosen foreign forum), so long as those expenses were (if submission were the limiting technique) incurred before Y's submission or (if confining Y to his reliance measure were the technique) incurred in reliance on the EJC being performed.

VI. CHOICE OF THIRD FORUM

This is a comparatively rare occurrence, but suppose X brings an action against Y in foreign forum 1, in breach of an EJC choosing the exclusive jurisdiction of foreign forum 2—can Y then maintain an action for breach of that foreign EJC in England? He might want to do this because X has assets in England.

[131] This assumes that the foreign EJC can be interpreted (pursuant to the relevant governing law) to cover satellite litigation such as actions for damages for breach of the EJC.

It seems that an anti-suit injunction would only be granted by an English court if England was, notwithstanding the foreign EJC, at least a natural forum for the dispute.[132] Given that an action for damages is arguably more efficacious than an anti-suit injunction, it is submitted that at least the same considerations should apply to Y's right to damages. Moreover, if proceedings were brought for breach of the EJC, then, provided the foreign EJC could be interpreted (pursuant to its relevant governing law) as extending to litigation arising under it, X would be entitled to a stay unless 'strong reasons' to the contrary were shown. Hence there would be no need to resort to any limiting factors on Y's domestic right to damages; the standard English rules of jurisdiction will resolve the matter satisfactorily.

VII. CONCLUSION

When X sues Y in breach of an EJC, Y can certainly recover those costs and expenses he incurs in pre-emptive action brought to resist that action in the foreign forum. Y will not be able to recover costs and expenses incurred in pre-emptive action in England except to the extent Y is awarded them by the court as costs as a result, it seems, of the *Quartz Hill* principle. Although problems of quantification should not debar Y's right to damages, there are clearly concerns which could best be articulated as concerns of potential abuse of process or infringement of comity. Before such concerns should be allowed to limit the domestic law right to damages, however, a technique drawn from the domestic law, such as submission or waiver, needs to be employed. Application of such a technique would preclude Y unwinding any substantive award made by a non-contractual forum and will also preclude Y from recovering any excess litigation costs in that substantive hearing which he would not have incurred had he been sued in the chosen forum.

To the extent that a substantive award is made by an English court after it decides not to stay proceedings despite a foreign EJC, another potential reason for not allowing Y to unwind that award is practical inconsistency. Similarly, to the extent that excess costs were incurred in substantive proceedings brought in England in breach of a foreign EJC, the *Quartz Hill* principle constitutes another reason for not allowing Y to recover those excess costs.

Damages for breach of EJCs are potentially far-reaching, but that breadth brings with it difficulties. As foreshadowed, there are no neat, universal solutions to these difficulties. All that can be hoped is that identification of a limited number of principles will assist in resolving particular situations when they do (as they are very likely to) arise in the near future.

[132] See *Airbus Industrie GIE v Patel* [1999] 1 AC 119, at 138–9, although not strictly speaking considering cases where there is an EJC. See *Dicey & Morris Conflict of Laws*, (London, Sweet at Maxwell) at p 419.

15

Interpreting Employment Contracts: Judges, Employers and Workers

SIMON DEAKIN

I. INTRODUCTION

THE AIM OF this chapter is to compare prevailing judicial interpretations of employment contracts with those of employers and workers. It focuses on a central issue in contemporary employment law, namely the construction of contractual documents and working arrangements for the purposes of determining whether an individual worker is an employee or self-employed. Juridical analysis turns, formally, on the test of 'mutuality of obligation': do the parties to the contract make mutual commitments to make work available (on the part of the employer) and to be available for work (on the part of the employee)? On the face of it, this is an approach to construction which is based on principles of general contract law, and as such, relatively uncontroversial. However, we do not have to dig far below the surface of the mutuality test to see that the test is founded on assumptions about choice, control and risk: *choice*, because the content of the legal relationship is assumed to have been determined by the free will of the individuals concerned; *control*, because the courts continue to understand the essence of the personal employment relationship in terms of the employer's power to direct the employee; and *risk*, because the parties' choice of employment arrangement is seen to embody a particular combination of protections and liabilities. These assumptions are deeply rooted in juridical discourse and can be traced back to earlier tests used by the courts to define employment status.

Recent empirical research in which the author was involved provides an opportunity to compare the prevailing juridical view of employment with the perceptions of employers and workers. Research was carried out for the Department of Trade and Industry in 1998–99 on how far the current tests of employment status were creating uncertainty in practice.[1] This took the form of a large-scale survey designed to estimate the numbers potentially excluded from certain forms of employment protection, and case studies of the current practice of employment contracting. The evidence collected included standard form employment agreements and accounts by workers of their experience of casual and flexible forms of work. Two conclusions stand out from this research. First, on the basis of survey evidence, there is a large proportion of the labour force which does not fit neatly into the two categories of 'employed' and 'self-employed' which the law uses to classify work relationships. Secondly, the case studies suggest that there is a considerable disjuncture between the assumptions of choice, control and risk which underlie the legal tests, and the perception of these issues by workers whose employment status is most in doubt.

The chapter is developed as follows. Section II below discusses the legal tests used to determine employment status and the role within them of contractual interpretation. Section III outlines the relevant empirical evidence. Section IV argues, on the basis of the previous two sections, that the law is currently operating dysfunctionally, and considers two possible reforms aimed at enhancing its effectiveness: the use of a different definitional test, based on the 'worker' notion, and the application to standard-form employment contracts of the techniques of legal regulation adopted in consumer contracts. It concludes that while each of these may prove useful in practice, there is a danger that, notwithstanding statutory reform, the assumptions underlying the courts' current approach to interpretation are likely to reappear in a new guise.

II. THE ROLE OF CONTRACTUAL INTERPRETATION IN DETERMINING EMPLOYMENT STATUS AND STATUTORY EMPLOYMENT RIGHTS

A. The Mutuality Test

Currently, most rights under the main labour law statutes—the Employment Rights Act 1996 (ERA) and the Trade Union and Labour Relations (Consolidation) Act 1992 (TULRCA)—apply only to employees. For this purpose, an employee means 'an individual who has entered into or works under (or, where the employment has ceased, worked under) a contract of employment' and a contract of employment means 'a contract of service or apprenticeship, whether express or implied, and (if it is express) whether oral or in writing'.[2] The substantive meaning of the contract of

[1] See B Burchell, S Deakin and S Honey, *The Employment Status of Individuals in Non-Standard Employment,* EMAR paper no 6 (London: DTI, 1999), http://www.dti.gov.uk/er/emar/emar6.pdf.
[2] Employment Rights Act [ERA] 1996, s 230(1), (3).

Table 1. The relationship between factors and tests for classifying employment relationships

Control	duty to obey orders
	discretion on hours of work
	supervision of mode of working
Integration	disciplinary/grievance procedure
	inclusion in occupational benefit schemes
Economic reality	method of payment
	freedom to hire others
	providing own equipment
	investing in own business
	method of payment of tax and NI
	coverage of sick pay, holiday pay
Mutuality of obligation	duration of employment
	regularity of employment
	right to refuse work
	custom in the trade

(Source: Burchell, Deakin and Honey, *The Employment Status of Individuals in Non-Standard Employment* (1999)).

employment is not made clear by statute; it has been left up to the courts to decide this question, applying common law tests. The four tests which have become best known are 'control', 'integration', 'business reality', and 'mutuality of obligation'.[3] Behind each of these tests lies a set of 'factors', such as the method of payment chosen by the parties, the length and stability of the employment relationship, and the degree of coverage of disciplinary and grievance procedures (see Table 1). The weight which courts attach to any particular factor appears to be a matter of discretion, in part because lower courts are only subject to review if they commit errors of law in their identification of the relevant tests or in their application.[4]

Although the other three tests are still referred to and are by no means irrelevant, the major focus in applying the common law criteria of employee status since the late 1970s has been on the 'mutuality of obligation' test. Where the existence of mutual obligations to provide work (in the case of the employer) and to accept any work which is offered (in the case of the worker) is in doubt, the relationship may be classified as one of self-employment, or otherwise outside the scope of the 'employee' concept. These exclusionary aspects of the mutuality test adversely affect homeworkers,[5]

[3] See generally S Deakin and G Morris, *Labour Law* 3rd edn (London, 2001), pp 146–168.
[4] See *O'Kelly v Trusthouse Forte plc* [1983] IRLR 396.
[5] *Airfix Footwear Ltd v Cope* [1978] ICR 1210; *Nethermere (St Neots) Ltd v Taverna and Gardiner* [1984] IRLR 240.

agency workers,[6] 'zero-hours' contract workers[7] and workers in casualised trades or occupations.[8]

Moreover, because of the way in which the rules of continuity of employment operate,[9] it is often necessary for an individual to establish not simply that he or she was employed under a contract of employment at some point, but that the contract remained in force for a sufficient period of time for them to acquire the necessary statutory continuity. Although the continuity rules are more straightforward to apply than used to be the case, following the abolition of the 8-hour and 16-hour thresholds in the mid-1990s,[10] establishing that a contract of employment exists during periods in between separate jobs remains problematic. Statute itself only allows for a few exceptional periods of non-employment to be counted towards continuity; otherwise, a contract must be implied.

For 'zero hours contract' workers and others whose working patterns are irregular or interrupted, this is a major problem. Hence in *Carmichael v National Power plc*,[11] the issue was whether the applicants, who worked periodically as tour guides at a power station, had contracts of employment between ad hoc hirings; the House of Lords, restoring the ruling of the employment tribunal, ruled that they did not. This meant that they could not claim the right to receive a written statement of particulars of employment from their employer, for which one month of continuous employment is required.

It is not a straightforward matter in general to determine how far the parties to an employment relationship are free to determine the status of the supplier of labour. On the one hand, the courts have said many times that they will disregard a 'label' attached to the relationship by the parties. Hence, if other considerations (of the kind considered above) clearly point towards employee status, an agreement between the parties to the effect that the individual is self-employed will have no legal effect. The 'relabelling clause' will only carry much weight if other factors do not clearly point in one direction or another.[12]

Legislation governing waivers is also relevant here. Any agreement by the individual to waive his or her protective rights under ERA 1996 or

[6] *Wickens v Champion Employment Agency* [1984] ICR 365; *Ironmonger v Movefield Ltd* [1988] IRLR 461; *Pertemps Group plc v Nixon,* 1 July 1993, unreported, EAT/496/91.

[7] *Clark v Oxfordshire Health Authority* [1998] IRLR 125.

[8] *O'Kelly v Trusthouse Forte plc* [1983] IRLR 369; *Carmichael v National Power plc* [1998] IRLR 301, [2000] IRLR 43.

[9] For the source of the rules governing continuity of employment, see ERA 1996, Part XIV, ch 1; and for discussion of the relationship between contract and continuity, Deakin and Morris, *op cit*, at pp 203–208.

[10] By 1995/1. See generally Deakin and Morris, *op cit*, at pp 193–198.

[11] [2000] IRLR 46.

[12] See *Cataraman Cruisers Ltd v Williams* [1994] IRLR 386.

TULRCA 1992 is void.[13] On the other hand, there is an unfortunate element of circularity about arguments from the prohibition on waivers—the anti-waiver law is only effective if the individual comes under the protection of the statute, but that is the very issue at stake when the court or tribunal is deciding issues of employment status.[14]

A more intractable problem is that, for all their talk of disregarding 'labels', the courts have also reiterated that there is nothing to prevent the parties *voluntarily* accepting an arrangement which, *objectively speaking*, is one of self-employment: '[a] man [sic] is without question free under the law to contract to carry out certain work for another without entering into a contract of service. Public policy has nothing to say either way.'[15] In principle, the form in which labour is contracted is a matter for the parties themselves and not for the law to decide. In practice what this means is that the issue will be settled according to the approach taken by a particular court or tribunal towards construction of the contractual arrangements in question, without regard to considerations of public policy concerning the desirability of ensuring a clear and uniform application of protective legislation. Those few cases in which the courts have said that they will take notice of this public policy point all concern the interpretation of the protective statutes themselves—not the construction of 'private' contractual arrangements.[16]

B. Decoding 'Mutuality of Obligation'

Contractual waivers and exclusions aimed at denying employee status or, in situations where this status is not in doubt, placing obstacles in the way of employees achieving the necessary continuity of employment to qualify for protective rights, can only work if the courts take a particular approach towards the interpretation of employment agreements. This is that contracts made between workers and employers are agreements made by the individual parties at arms' length, and which therefore fall to be interpreted according to normal canons of contractual construction. That this is the prevailing approach of the courts is evident from numerous decisions.

[13] See ERA 1996, s 203; TULRCA, s 288.
[14] See Deakin and Morris, *op cit*, at pp 152–154.
[15] *Calder v H Kitson Vickers & Sons (Engineers) Ltd* [1988] ICR 232, 250 (Ralph Gibson LJ).
[16] See *Jones v Tower Boot* [1997] IRLR 168, 171–172; *Harrods Ltd v Remick* [1997] IRLR 583, 583; *MHC Consulting Ltd v Tansell* [1999] IRLR 677, 679, and on appeal sub nom *Abbey Life Assurance Co v Tansell* [2000] IRLR 387, 390. Significantly, these were all decisions under the anti-discrimination Acts, where considerations of human rights at work appear to be more in the forefront of the judges' minds than is the case when they are deciding issues of employment protection law.

It does not take very much imagination to see that many of the clauses inserted into standard form agreements were put there by the employer to deflect employee status. This is the case, for example, with clauses which purport to grant the worker the right to appoint a substitute to take their place if they are unable to work (which we may call a 'substitution clause'). If read literally, this would amount to a denial not just of employee status (since mutuality of obligation would be lacking) but, going further, of the essential element of a personal commitment to provide labour which is also present in the contract for services. Another technique is to assert that there is no obligation for the employer to provide work, nor for the employee to accept it (let us call this a 'no mutuality clause'). If, in practice, there is evidence that work is carried out continuously, and that the worker does not regard him or herself as having discretion to take time off or to appoint a substitute whenever they feel like it, it should be open to a court to treat the clause as nothing more than the boilerplate which it so clearly is.

But in several recent decisions, courts have taken these clauses at face value. The trend began with a 'substitution clause' in *Express and Echo Publications Ltd v Tanton*[17]and continued with a 'no mutuality clause' in *Stevedoring & Haulage Services Ltd v Fuller*.[18] There, the employment tribunal had taken the view that it could look to the practice of employment in order to establish what the parties must have intended the terms of the contract to be. On that basis, it implied an agreement between the two parties which would have satisfied the mutuality test. This approach, while admittedly heterodox from the viewpoint of general contractual construction, has highly respectable antecedents in employment law.[19] Yet the tribunal's decision was reversed, the Employment Appeal Tribunal ruling that it was not possible to have regard to the practice or conduct of the employment relationship where the wording of the express contract was clear. Behind the mutuality test, then, is the assumption we may refer to as *consent*: the courts take the view that an express agreement, even one which is plainly based on a standard form proffered by the employer, represents a *consensus ad idem* between the two parties.

There are other more or less tacit assumptions at work here. In dealing with cases of casual workers, the courts are in effect using the contractual language of reciprocity to revive the old test of *control* in a new form. As we have seen, unless the worker agrees to be available for work on a continuing basis, and in this extended sense *at the disposal of the employer*, the agreement is said to lack the necessary mutuality of obligation. What counts here is not reciprocity in exchange as such, but a particular form of

[17] [1999] IRLR 367; cf *McFarlane v Glasgow City Council* [2001] IRLR 7 and *Byrne Bros v Baird* [2002] IRLR 96, discussed further, below s 2.3.
[18] [2001] IRLR 627.
[19] See in particular the judgments of Browne-Wilkinson J in *Jones v Associated Tunnelling Ltd* [1981] IRLR 477 and Lord Hoffmann in *Carmichael v National Power plc* [2000] IRLR 43.

mutuality under which the worker must cede autonomy to the employer over the timing and physical location of the work in order to count as an employee.

Thus in the employment sphere, 'mutuality of obligation' does not bear its normal meaning in contract law. A contract to work in return for pay will not lack 'mutuality' in the sense of the reciprocal promises needed for consideration as one of the elements in the formation of a contract. However, if the judgments in *O'Kelly v Trusthouse Forte plc*[20] are to be believed, such an arrangement, without more, will not be a contract *of employment*. The mutuality of obligation test only *looks as if* it has a connection to mainstream contract law. In truth, it is a test which emerged in the particular context of employment law, and which functions above all to exclude casual forms of work from the coverage of protective legislation. Although it can be traced back to nineteenth century decisions on the application of master and servant law (itself a further clue as to its true lineage),[21] the test is not mentioned in leading twentieth century decisions up to the late 1970s. Its emergence can be dated to the time[22] that modern employment protection legislation began to occupy a major place in the practice of employee relations, and when the difficulty of fitting employment rights into the framework of casual work was becoming a live issue.

The third set of assumptions embedded in the mutuality test concerns *risk*. The parties to an employment agreement—whether it is a contract of employment or a contract for services—are understood to be opting into a schema of protections and liabilities, with one set traded off against the other. Thus in entering into a contract of employment, the employee trades off 'subordination' or acceptance of the employer's right to give instructions and to organise the carrying out of the work, in return for certain protections: at a basic level, a regular wage which a self-employed person would not expect to receive; at a more extended level, the full set of income and job security rights which go with continuous service under the current provisions of employment protection legislation. By contrast, an independent or autonomous worker employed under a contract for services cannot look to the employer for security of income or employment, but is able to take advantage of a more favourable tax and social security regime (in the sense that income tax is paid net of work-related expenses and National Insurance contributions are paid at a lower rate), as well as the autonomy of being able to decide when, where and how to work.

The risk perspective can be viewed either negatively or positively. In older authorities under the 'control' test, the courts treated the trade-off

[20] [1983] IRLR 369.
[21] See S Deakin, 'The contract of employment: a study in legal evolution' (2001) 11 *Historical Studies in Industrial Relations* 1, 20–21, for discussion of this case law.
[22] See in particular *Airfix Footwear Ltd v Cope* [1978] ICR 1210.

involved in employment sceptically, regarding the implicit loss of autonomy as inappropriate for workers who exercised discretion over the performance of tasks. Thus in *Simpson v Ebbw Vale Steel, Iron & Coal Co*, decided in 1905 at the dawn of modern social protection legislation, the Court of Appeal, when deciding that the Workmen's Compensation Act 1897 had no application in the case of a colliery manager who was killed in a mining accident, said of the legislation:

> It presupposes a position of dependence; it treats the class of workmen as being in a sense 'inopes consilii', and the Legislature does for them what they cannot do for themselves: it gives them a sort of State insurance, it being assumed that they are either not sufficiently intelligent or not sufficiently in funds to insure themselves. In no sense can such a principle extend to those who are earning good salaries.[23]

Almost a century later, we find the Employment Appeal Tribunal returning to the theme of dependence in *Byrne Bros v Baird* when attempting to determine the scope of the Working Time Regulations 1998, which impose maximum limits on weekly working hours:

> The reason why employees are thought to need such protection is that they are in a subordinate and dependent position vis-à-vis their employers: the purpose of the Regulations is to extend protection to workers who are, substantively and economically, in the same position. Thus the essence of the intended distinction must be between, on the one hand, workers whose degree of dependence is essentially the same as that of employees and, on the other, contractors who have a sufficiently arm's-length and independent position to be treated as being able to look after themselves in the relevant respects.[24]

In *Byrne Bros*, in contrast to its predecessor of 1905, the court of 2002 took a functional view of the legislation which was before it, and gave a broad interpretation to its scope.[25] In that sense, judicial attitudes towards social legislation have clearly changed in the course of the past one hundred years. Yet the two judgments have one crucial factor in common: the idea that it is possible to distinguish clearly between two different groups, on the one hand those subject to a regime of managerial coordination in return for protection from the risks of employment, and on the other those whose autonomy (in either a personal or an economic sense) takes them outside the coverage of social protection. In this respect, the contemporary tests of employment status continue to reflect the notion of a 'binary divide'[26]

[23] [1905] 1 KB 453, 458 (Collins MR).
[24] [2002] IRLR 96.
[25] The case turned on the meaning of the 'worker' concept: see the discussion in the next subsection of the text.
[26] See M Freedland, 'The role of the contract of employment in modern labour law', in L Betten (ed) *The Employment Contract in Transforming Labour Relations* (Deventer: Kluwer 1995).

between employment and self-employment which is at the foundation of modern legislation on employment protection, social insurance and income taxation.

C. Beyond Mutuality? The 'Worker' Concept and the Regulation of Agency Work

Dissatisfaction with the mutuality test has been one of the moving forces behind the recent attempt to extend the scope of employment legislation through the use of the concept of the *worker.* Under the Employment Rights Act 1996, section 230(3), a worker is defined as 'an individual who has entered into or works under (or, where the employment has ceased, worked under)—(a) a contract of employment, or (b) any other contract, whether express or implied and (if it is express) whether oral or in writing, whereby the individual undertakes to do or perform personally any work or services for another party to the contract whose status is not by virtue of the contract that of a client or customer of any profession or business carried on by the individual'.[27] Similar definitions have been used in the context of recent legislation on the national minimum wage[28] and the organisation of working time.[29]

In *Byrne Bros v Baird*,[30] the leading case to date on the 'worker' definition, the Employment Appeal Tribunal (EAT) took a policy-orientated view of the statutory concept, holding that it was intended 'to create an intermediate class of protected worker, who is on the one hand not an employee but on the other hand cannot in some narrower sense be regarded as carrying on a business.' At the same time, the Court recognised that the 'wording of limb (b) [that is, the extended 'worker' definition] gives no real help on what are the criteria for carrying on a business undertaking in the sense intended by the Regulations—given that they cannot be the same as the criteria for distinguishing employment from self-employment'. The EAT then went on to say this about the distinction between a worker and an independent contractor falling outside the scope of protection altogether:

> Drawing that distinction in any particular case will involve all or most of the same considerations as arise in drawing the distinction between a contract of service and a contract for services—but with the boundary pushed further in the putative worker's favour. It may, for example, be relevant to assess the degree of control exercised by the putative employer, the exclusivity of the engagement and its typical duration, the method of payment, what equipment

[27] See also TULRCA 1992, s 296(1).
[28] National Minimum Wage Act 1998, s 54.
[29] Working Time Regulations 1998, reg 2.
[30] [2002] IRLR 96.

> the putative worker supplies, the level of risk undertaken etc. The basic effect of limb (b) is, so to speak, to lower the pass-mark, so that cases which failed to reach the mark necessary to qualify for protection as employees might nevertheless do so as workers.

This is a frank acknowledgement that the test for determining 'worker' status may well not be fundamentally different from that which is applied to the 'employee'—if there is a difference it is one of degree, not kind. The outcome in this case, in which the court found that the applicants were 'workers' and rejected an argument that a 'substitution clause' in their contracts prevented this outcome, suggests that the introduction of the 'worker' concept may just tip the balance in certain cases. But if the dictum just quoted is any guide, it is highly likely that the same assumptions of choice, control and risk which underlie the court's approach to the 'employee' issue will continue to be relevant in the changed statutory context of the 'worker' test.

Agency workers are also the focus of specific statutory intervention. Under legislation regulating employment agencies and businesses, they are regarded as akin to employees for the purposes of tax and national insurance contributions.[31] However, this legislation does not stipulate what their status should be for other purposes, including employment protection. At common law, it is sometimes said that their contractual position is unique, and falls outside the normal classifications,[32] but this unduly neglects the possibility that they may be employees of either the agency or the user. The normal approach is to regard them as contracted to the agency. However, a number of decisions suggest that unless they are guaranteed continuous work by the agency (which is unlikely), they will lack mutuality of obligation and hence be employed under contracts for services.[33] In relation to the user, the mere presence of control, in the sense of managerial coordination, is not sufficient to give rise to a contract of employment, in the absence of additional evidence of a contractual nexus.[34] However, both the user and the agency may be required to treat agency workers equally with respect to sex, race and disability discrimination.[35] In addition, agency workers have

[31] Employment Agencies Act 1973 and Conduct of Employment Agencies and Employment Businesses Regulations 1976, SI 1976/715.

[32] *Construction Industry Training Board v Labour Force Ltd* [1970] 3 All ER 220.

[33] *Wickens v Champion Employment Agency Ltd* [1984] ICR 365; *Ironmonger v Movefield Ltd* [1988] IRLR 461; *Montgomery v Johnson Underwood* [2001] ICR 819.

[34] *Hewlett Packard Ltd v O'Murphy* [2002] IRLR 4.

[35] See, in the case of the user or 'principal', Sex Discrimination Act 1976, s 5, Race Relations Act 1976, s 7, Disability Discrimination Act 1995, s 12, and *Abbey Life Assurance Co Ltd v Tansell* [2000] IRLR 387; in relation to the agency, see Sex Discrimination Act 1975, s 15, Race Relations Act 1976, s 14, and Disability Discrimination Act 1995, s 68(1). It is far from clear whether the user is required to maintain the principle of equal treatment of agency workers with regard to other, directly employed workers in that employment unit: see *Allonby v Accrington and Rossendale College* [2001] IRLR 364.

been brought under the scope of minimum wage and working time legislation by imposing the relevant obligation on either the agency or the user, depending, first, on which of the two of them is contracted to pay them and, failing that, on which one actually does pay them.[36]

The problem posed by agency work is that the idea of the trade off of 'subordination' for 'security' is undermined by the form of the contractual relationship: the 'coordination' function vests in the user while a residual 'risk' function is left with the agency. The resulting problems of attribution of responsibility can be overcome only by legislative intervention, but this is complex in its effects, and cannot necessarily deal in advance with all the problems which may arise. In truth, the attitude of the UK legislature to agency work is ambiguous. It falls short of installing the kind of protective regime which would ensure, through a combination of independent and concurrent liabilities, that either the user or the agency, or both together, assume the normal responsibilities of the single employer. It fails to take this step because to do so would no doubt operate as a deterrent to the use of agency work in some cases. However, the question is whether it is *legitimate* to make use of the agency option in cases where the principal reason for doing so is simply the greater opportunity it offers, by comparison with direct employment, to offload risks on to the individual worker, rather than greater flexibility in the organisation of work.

III. EMPIRICAL EVIDENCE

A. The Empirical Project

We turn now to the findings of the empirical research. The principal objectives of the research carried out for the DTI were to estimate the number of individuals who might be affected by the wider adoption of the concept of 'worker' in employment law and to identify the sources of uncertainty in the application of the current legal tests of employment status. To tackle these research questions required a multi-method approach, involving a large, representative sample of individual respondents, and detailed, qualitative studies of employment practice in a small number of cases of flexible or 'non-standard' work.

The first wave of data was collected from a representative sample of 4,006 members of the British workforce who were interviewed in January and February 1998. After an initial screening to exclude those who were unambiguously employees or self-employed, 1,182 were asked a range of questions about the nature of their employment relationships. The aim here was

[36] National Minimum Wage Act 1998, s 34; Working Time Regulations 1998, reg 36.

to go beyond the 'self reporting' method which is used to assess the number of employees and self-employed in the quarterly Labour Force Survey (LFS). The LFS relies entirely on respondents' own perceptions of their employment status. In practice, as many of the cases which reach the courts only too clearly indicate, an individuals' own understanding of his or her status may diverge significantly from that of a court or tribunal.

In an attempt to get round this problem, the questionnaire was drafted in such a way as to reproduce, through specific questions, the kind of issues which a court or tribunal would take into account. Individual responses were then analysed to see how far respondents could be categorised as *unambiguously* employees or self-employed. Respondents were regarded as 'clearly employees' if they defined themselves as such; were paid a salary or wage; held what they regarded as a permanent job; *and* had no non-standard working patterns (such as fixed-term, casual or part-time work). Conversely, they were 'clearly self-employed' if they were either a director or partner in their own business, *and/or* employed others. On this basis, 64 per cent of all respondents were classified as clearly employees, and 5 per cent were clearly self-employed. This compares to the 86 per cent of the sample who saw themselves as employees, and 13 per cent who saw themselves as self-employed (a result which corresponds to that of the LFS for the same period). The first result of the empirical research, then, was to show that a large proportion of the national labour force, 30 per cent on this estimate, was employed under terms and conditions which created some degree of uncertainty over their employment status.

The next step was to try to arrive at an estimate of how many individuals could be classified as 'workers' under the extended definition of that concept. For this purpose, the analysis focused on the group of individuals which would remain outside the 'worker' definition, that is to say, the 'independent self-employed'. These were identified as those respondents who had worked for more than one employer in the six months prior to the questionnaire, and who 'passed' the tests of 'economic reality' for autonomous work: in other words, they were able to sub-contract, they were not paid a wage or salary, they paid their own income tax and national insurance contributions, and they were not entitled to receive either sick pay or paid holidays. This group constituted 8 per cent of the total sample, leaving a figure of 92 per cent of the labour force as 'workers'. But of this 92 per cent, many would be affected by uncertainty as to their status, because they, in turn, would only 'fail' perhaps one or two of the economic reality tests. When this factor was taken into account, the proportion of the total labour force which could be classified as *clearly* in the 'worker' category fell to only 80 per cent.[37]

[37] For the precise basis of this calculation, and an explanation and justification of the methodology used to arrive at it, see Burchell, Deakin and Honey, *op cit*, at pp 43–46.

The message coming out of the quantitative phase of the study is therefore the following: whichever concept is used, 'employee' or 'worker', the nature of modern employment relations is such that there will inevitably be a large group of individuals whose employment status is in an unclear 'grey zone' between employment and self-employment. The 'worker' concept, in so far as it can be assumed to map reasonably closely on to the test of 'economic reality', reduces the size of the 'grey zone', but does not eliminate it.

The second wave of data collection involved re-visiting a small proportion of that part of the sample whose employment status was uncertain or ambiguous in order to gather more in-depth information in less structured discussions, and to analyse their written contracts of employment or terms and conditions of employment. These qualitative case-studies were based on a combination of focus groups and individual semi-structured interviews, involving altogether 36 respondents. The interviews were carried out in the spring and summer of 1998. Documentary evidence, including individuals' contracts of employment and written statements of terms and conditions, was also obtained.

A full analysis of all the cases of uncertain employment status in the qualitative wave of the research has been carried out elsewhere.[38] What follows is an analysis of how the issues of *choice*, *control* and *risk* were viewed by those who were interviewed.

B. Choice: The Use of Standard Form Contracts to Deflect Employee Status

The case studies provide abundant examples of employers using standard-term employment agreements as devices to insert terms which have the effect of altering employment status. Thus it was common to find that contracts entered into by agency workers with their agency contained both a 'relabelling' clause and a 'no mutuality clause'. In one case, the relabelling clause stated that,

> for the avoidance of doubt, these terms shall not give rise to a contract of employment … and therefore the [worker] will not have the statutory rights accorded to employees.

This was unlikely to have negatived employee status since it simply represented the view of one of the parties as to the agreement's legal effect. The 'no mutuality' term stated:

> The [worker] acknowledges that it is the nature of temporary work that there may be periods when no suitable work is available and agrees (a) that suitability

[38] Burchell, Deakin and Honey, *op cit*, chs 5 and 6.

> shall be determined solely by the [agency and user] and (b) that the [agency and user] shall incur no liability towards the [worker] should they fail to offer opportunities to work … .

This term stands a much greater chance of excluding employee status, so long as the court takes it at face value, and does not permit evidence of any contrary practice of regular work to be set up against it.

In another case, several specimen contracts with agencies were obtained. The specimen contracts all contained clauses which appeared to be aiming at the exclusion of employee status by a variety of means. One was to deny that the arrangement constituted a legally binding contract of employment between the worker and either the agency or the user (instead the arrangement was stated to be a contract for services); another was to stipulate that the worker was not required to accept any assignment, nor was the agency required to find an assignment for the worker (the mutuality of obligation point). Similar techniques can be seen in the contractual documentation analysed by the Employment Appeal Tribunal in the *Pertemps* case.[39] In that case, the EAT considered that the terms of the formal written contract were an important indicator that the agency worker concerned was not an employee.

The contractual documents of agency workers also indicated that once each assignment was accepted by the worker, the worker came under a number of stated obligations, for example to cooperate with the user's staff and to follow the work rules of the user. There are dicta to suggest that making such obligations explicit could be seen as evidence that the relationship with the agency was *not* based on a contract of employment, for the very reason that such obligations are normally implied into such a contract; if the worker were truly an employee, it may be argued, there would be no need to spell them out.[40] An alternative view is that there is nothing to prevent an employer (or any other party) seeking to make explicit what would otherwise be implicit. On this basis it could be said that spelling out the obligations of obedience and care does not help to settle the issue of status either way.[41] In the present, confused state of the law, however, it is not surprising that employers seek to take advantage of judicial dicta which open the way to a finding that the worker is self-employed.

In another case, the respondent worked as a freelance copy-editor for a publishing house. Under a highly detailed and lengthy contract issued by the publishing house, a clear distinction was drawn between 'freelancers' and 'employees': the documentation referred to the importance of this distinction from the point of view of complying with Inland Revenue and

[39] Unreported, 1 July 1993, Appeal No EAT/496/91.
[40] This was the view of the EAT in the *Pertemps* case, *ibid*.
[41] This was the approach taken by the Court of Appeal in *McMeechan v Secretary of State for Employment* [1997] IRLR 353.

Contributions Agency rules on tax and national insurance contributions. The documentation spelt out the *employer's* view that the distinction between freelancers and employees was part of what made freelance work attractive to both sides, and specified reasons for the view that the respondent was not employed as its employee: as a freelance she would not normally work on the employer's premises; she controlled when the work was done, within agreed completion dates; the contractual relationship between the employer and the respondent only began when an order was placed and ended when the employer paid for the completed work; the respondent did not have an email, telephone number or desk at the employer's place of work; and she was paid only on presentation of an invoice for work done. The *respondent's* account of her working relationship, given during her interview, suggests that she was uncertain as to its effects. When asked whether the employer had a duty to provide her with work she replied 'absolutely', suggesting that she saw herself as having what a lawyer would refer to as 'mutuality of obligation'. However, on the basis of her agreement, she concluded that she did not have a continuing or 'global' contractual relationship with the employer. She also stated in her initial survey interview that she could if necessary hire others to work with her, which would count in favour of her being independently self-employed.

Overall, the case studies suggest that employers make frequent use of standard form contractual documentation which is drafted with the case-law on employee status in mind. These documents are meant to shield employers from legal liability by providing evidence which can rebut a claim of employee status. This type of documentation is also used as a signal to workers that they have self-employed status. The contents of documents may bear little relationship to the *practice* of a particular employment relationship; indeed, they may contradict it. There is very little evidence that either employers or employees see this type of documentation in the way that it is viewed by the courts, namely as embodying a contractual *consensus ad idem*.

C. Control: Perceptions of Managerial Power

The juridical divide between employment and self-employment is founded on an assumption that employees are subject to a degree of managerial coordination or control which does not apply to independent contractors. However, the case studies suggest that, in practice, perceptions of control and accountability are much more complex than a simple division between employment and self-employment would imply. Self-employed respondents commented that they came under pressure to accept work from particular clients and had to operate to very tight deadlines. Similarly, respondents employed on zero-hours or on-call contracts saw themselves as being required to respond to the employer's demands, even if their contracts suggested that they had the right to turn down work offered to them. A freelancer who had no formal commitment to a regular client, and had no formal contractual expectation of receiving continuous work, commented that in practice it was often difficult to refuse work:

> *So what you're basically saying is you are free to decide when you work and for whom or are you not free to decide when you work?* Well, I'm free inasmuch as I'm not free—well, to be honest, no—I'm free to decide when I work, I'm not free to decide who I work for at the moment because I have to take whatever I can get. I always pretty much have had to take whatever I can get. So I'm not free to decide who I work for.

In turn, very long hours, unsocial hours working and variability of working hours were problems both for employees and for the self-employed. For certain employees, including managerial and professional workers, hours were perceived as being flexible in the employer's favour. A charity manager said that while her contracted hours were 37 a week, she normally worked 60 or more hours, with adverse consequences for her health:

> Yes, its 37 hours in theory and flexible. If I do evening and weekends, in theory I have time off in lieu. One of the reasons I am not well now is that I have been overdoing it over 2 years and I look around the voluntary sector and see a lot of people under stress and getting ill because they are trying so hard to make ends meet and do the job with limited resources and be everything to everybody and fulfil bottomless demands—so you end up doing excess hours.

Some agency workers considered that their status was a flexible one which gave them the option of refusing work. However, there were also cases in which respondents employed through agencies worked for long periods alongside permanent staff doing very much the same work. Here, there was no sense in which agency employment gave the worker greater 'control' over his own working arrangements:

> *Do you look upon yourself as being basically your own boss, so that you can decide whether you want to stay with them or not?* No, I wouldn't say my own boss because you know you are working for a firm—if it wasn't them, it would be another agency. *Would you rather have a job with a firm or are you quite happy working for an agency?* I would prefer a permanent job, yes, if [the client] said OK we'll take you on permanently, I would say yes. *What would the advantages be?* There would be a more stable sort of environment and you would also have a pension scheme and … the benefits that go with [being with] a company … .

In short, the predominant view among those interviewed was that under non-standard working arrangements, control over when and how work was done very often vested with the employer or client, regardless of the legal form of the relationship in question.

D. *Risk*: The Flexibility-Security 'Trade off' and the Experience of Non-Standard Work

As we have seen, the approach of the courts is based on the assumption that individuals trade off security for autonomy and flexibility when opting for non-standard work or self-employment in place of employment as an employee. This found an echo in the comments made by some respondents. Some explicitly identified with the idea of a 'trade off' between the stability and security of employee status, on the one hand, and the greater autonomy of self-employment, on the other. Freelance workers and the self-employed contrasted the 'control' to which they had had to submit when they were employees, with the greater freedom but also the responsibility which being self-employed entailed:

> The advantages of being employed over self-employed are obvious in that you've got a regular income coming in and the buck doesn't stop with you! Although I've always had fairly senior positions, there was always somebody who I could offload on to, if you like. And so there were advantages in that. I was made redundant once when the company closed down ... it was completely out of my hands and I don't like that. I like to be in control, so the advantage of being self-employed there—I was in control. And of course, I've worked for large companies and small companies (child minder).
>
> The only good thing about working for an employer is that you don't take any of the strain, you start work at 8.30 and you finish at 5.30—you don't have to worry about what happens overnight. Here, I sort of think "'Oh, you know, who's got the keys to the next venue'" because we swap around, and like today, I was worried about [an employee] because she was very quiet this morning and I thought "'Oh, I've upset her'" (health promotion worker).

The downside of being self-employed was, according to one respondent, 'being poor' but the advantages included: 'you do the job and as long as you're doing the job, people don't worry about what you're doing at that precise moment they walk past your desk'.

Freelance workers appeared to value the autonomy to arrange a pattern of working which suited their needs:

> I find that I have developed a way of working, over a series of fifty or so contracts, that I feel very happy with. I wasn't given guidance, but I didn't find that a problem .... I could work out the best way to do things and interact with the office .... I don't even have to tell them [clients] really. I just take the

> product that they give me, and then return it to them and hopefully that's what is required, and then I'm done (freelance editor).
>
> I like the fact that I'm my own boss. To a certain extent, you set your own hours because if people come and they want you to start at 6 in the morning if they're nurses or they do shift work, you can always say no, you don't work those hours. It's nice to be at home in the summer. It's nice to call the tune, basically (child minder).

At the same time, the case studies suggest that choice is conditioned by factors which include the need to fit in with family arrangements; the cost of retraining following time spent out of full-time work; the time and complexity of setting up a business; discrimination against older workers; and the unavailability of regular work. The need to fit working hours around childcare requirements was one of the principal factors which motivated those who had taken up childminding:

> I had a good job once! I started doing this when I had my own children, you see, and it's difficult when you've got children. I didn't want to put mine with a childminder, so you incorporate an extra one into the home and when your children are at school, you need a job that will put you here when they're here and in school holidays.

Age discrimination was also cited as a factor in the context of a return to full-time employment after a break for child-rearing. The difficulty of getting a regular job made agency work attractive for one individual in this position:

> Initially I did hope to get another permanent job, either full-time or part-time, but I knew my age was against me. I really think equal opportunities is a name only ... that's with age, disablement. Yes, they'll call you for an interview because there are equal opportunities. But you know full well that you are not going to get it. So I am quite content with now, the way I am working and making a life for myself and paid of course (agency worker).

This respondent was in her late 50s and had returned to work after a break of almost 20 years raising her children. She was also registered as disabled.

In addition, many respondents saw non-standard work as inferior to regular working arrangements. Agency workers and fixed-term contract workers, in some cases, took these forms of work because permanent work was not made available to them. One respondent had been dismissed by his employer and re-employed on a self-employed basis. Once this happened he decided to set up his own business, but had not previously considered doing so.

Moreover, on the balance of evidence of the interviews there was no straightforward division between the perceptions of employees, who might have been expected to have feelings of security, and those of the self-employed, who might have been expected to accept insecurity in return for the prospect of greater reward. Many employees had concerns about the inherently insecure nature of their jobs, concerns which were largely outside their control. This was particularly the case with fixed-term contract workers. A manager of a charity commented on

> the anxiety and insecurity of it all and I can keep telling myself, well, in the commercial sector this is the fact of life, if you don't sell what you do, you don't survive . . . . We have to sell what we are doing to fund us and if we don't do that successfully then we don't deserve to carry on.

A hospital doctor employed on a fixed-term contract as part of his training referred to a growing lack of security of employment:

> medicine's always been looked on as having a lot of job security, but I think as things are changing and the market environment's creeping in then that really is going.

This respondent had a three year contract, about which he said:

> It gives you a bit of stability but even three years isn't that long and after that, I may not be able to get a job at all. It's not common in medicine, but it does happen.

A further problem associated with employment on a series of fixed-term contracts was lack of clarity of the legal position.[42] A clerical worker in the public sector told us:

> I was a bit concerned but I didn't want to mention it because I didn't know whether my at the time my boss knew about the two year rule and I thought well, if he keeps me on longer, I'll keep quiet! But he knew—he obviously knew about it - because I saw a note from him asking the admin. person there—when is [X's] second anniversary, you know.)

Agency workers, likewise, had concerns with insecurity and also with difficulties of working alongside permanent staff who were paid more than they were:

> *Those sorts of people—the ones that don't like agency workers—have they indicated why they don't like them?* 'I think it was because they felt threatened first of all, because they were actually training us to do what they

[42] At the time of the study, an employee employed under a fixed-term contract could lawfully waive his or her entitlement to claim unfair dismissal or redundancy compensation upon the expiry of the contract (ERA 1996, s 197); in addition, an employee, whether employed under a fixed term or not, could not acquire general protection against unfair dismissal, or the right to redundancy compensation, until they had acquired two years of continuous employment.

> were doing and human nature being what it is, you're going to think—I'm going to train these, they're paying them less than me, I'm going to be out! I think once—they have had redundancies but it was voluntary and the people who wanted to go went. The other people are still there. And I think over the time of being there, their fears gradually eased off.

For some agency workers, difficult relations with the permanent staff were tempered by the feeling 'if you weren't doing this, you could be on the dole'. Even then, there was the possibility that the employment relationship could be brought to an end at very short notice; the same respondent described how:

> one Monday morning we were working in the city centre and as we went in there were two of our [agency] supervisors there and they had a list and saw who was coming in .... They were just literally stopping people at the door, sending them over, and I think in one go they got rid of about 150 people that morning.

Respondents on both sides of the employee/self-employed divide expressed concerns about health and safety. An agency worker reported difficulty in obtaining compensation after suffering an injury at work for which neither the agency nor the user would take responsibility. A self-employed construction worker commented on the high health and safety risks facing subcontractors on building sites:

> ... a lot of companies do this now. They put a scaffold up for brickwork or for people to put the roof on and then want you to go on the scaffolding after to put the window in, as soon as they've finished the roof ... because they've put the scaffolding up for the roofers, they've put a sign on it saying 'not suitable for anything else'. So if you go on after them and you fall off that scaffolding, it clears them of liability for your injuries. Which is common practice now.

A majority of the respondents also associated the work they were doing with the absence of long-term financial security. Several agency workers, casual and zero-hours contract workers, fixed-term contract workers and self-employed construction workers had no access to an employer's pension scheme and had made no provision for themselves. In turn, the downside of being an employee, however, was not simply a degree of inflexibility over working arrangements, but also, in many cases, the lack of any compensating security:

> As an employee you have set times and you are tied down to that job and in all honesty employers mess you about, because they know that if they were to get rid of you within a matter of days they could get someone else doing the same job. So they don't care for the worker like they should because they know there is someone there to replace them and that is one of the reasons why I set up on my own. I used to always get laid off at Christmas time. (self-employed construction worker).

In short, the qualitative wave provided rich evidence on individuals' reasons for choosing particular forms of work and their perceptions of flexibility and autonomy, on the one hand, and of insecurity and risk, on the other. Some respondents saw the advantages and disadvantages of particular forms of work in terms of trade-offs between flexibility and security, suggesting that they exercised a degree of choice in weighing up which form of work to adopt. In numerous cases, however, the choice of non-standard work was seen as influenced and constrained by external pressures, the most important of which were family commitments, retraining costs, age and disability discrimination, and the lack of availability of alternative work. In particular, for those with family obligations, it was a matter of necessity to find employment which offered them the opportunity to arrange their work around domestic commitments. Those returning to employment after a period of unemployment or after family commitments chose non-standard work because of the costs of acquiring or re-acquiring skills of the kind needed for a more stable and permanent position. There was a perception that it was easier for older workers to get employment with an agency than with an employer looking for a longer-term commitment.

Although many respondents clearly identified particular advantages and disadvantages with the form of work in which they were engaged, there was a blurring of the division between standard and non-standard work, and between employment and self-employment. Hence self-employment could result in a considerable restriction of personal autonomy in practice and to long and intense working hours, thanks to the need to meet tight deadlines and maintain reputation with clients. Employees in non-standard employment, conversely, commented on growing insecurity and stress caused by uncertainty over their future job prospects.

Both employees and the self-employed reported being affected in different ways by financial insecurity. Agency workers and the self-employed often had no access to pension schemes, and this was also a problem for employees working on fixed-term or task contracts. Many of the social and economic risks for which employment legislation makes provision were perceived as being common to both employment and self-employment; these include low pay, insecurity of work, health and safety risks and absence of long-term financial security.

IV. CONCLUSION: ALIGNING EMPLOYMENT LAW WITH EMPLOYMENT PRACTICE

Evidence of individuals' experience of non-standard work suggests that the legal division between employment and self-employment does not correspond to *social perceptions* of a clear divide between these different forms of work. In the context of these forms of non-standard work, then, the notions of choice, control, and risk which underlie the approach of the courts are significantly out of synch with the way employment relationships are viewed on the ground. This does not mean that the approach of the courts is, in itself, illegitimate. The tests used by the courts are the product of a long process of emergence which is largely internal to legal doctrine.[43] It does however suggest that employment law, at a social and economic level, is currently operating dysfunctionally.

One possible solution, which the legislature has adopted in recent years, is, as we have seen, to alter the basic definitional test for dependent labour, by substituting 'worker' for 'employee'. Whether this will be successful remains to be seen; as yet, there are too few decided cases for the issue to be clear. There is a danger that the same assumptions which shape judicial attitudes to the employment contract will simply re-emerge in a new form. Thus there may be certain types of work relationship which the courts will be reluctant to interpret as giving rise to protected status, whatever word is used to describe that status. If, however, the courts are prepared to use the criterion of 'economic reality' to shape the 'worker' concept, the empirical evidence obtained from the quantitative wave of the DTI study suggests that significantly fewer of those who currently think of themselves as 'employees' would be affected by uncertainty over their legal status. This would mark a significant change.

The main obstacle to aligning the courts' interpretation of employment contracts with those of the parties themselves is the rigid and artificial 'mutuality of obligation' test. Under these circumstances, progress in making employment law less dysfunctional depends upon the courts changing their approach to the construction of contractual documents and working arrangements. A first step (but only the first) would be to cease taking standard-form employment agreements at face value, and to consider applying to them the same kind of sceptical scrutiny which is applied to similar agreements in the consumer sphere. The courts cannot apply directly to employment contracts the provisions of the Unfair Terms in Consumer Contract Regulations 1999, nor does it seem that the Unfair Contract Terms Act 1977 will be of much use in the employment sphere.[44] But it is open to employment tribunals, as a matter of contractual construction, to have regard to the practice of employment as a guide to interpreting the terms of express agreements. It is surprising that they do not take this route more often, given recent dicta in the House of Lords encouraging them to do so.[45]

[43] See Deakin, 'The contract of employment: a study in legal evolution', *op cit*.

[44] For discussion of the reasons for the limited usefulness of the 1977 Act in this context following the judgment of the High Court in *Brigden v American Express Bank Ltd* [2000] IRLR 94, see Deakin and Morris, *op cit*, at p 269.

[45] In Lord Hoffmann's judgment in *Carmichael v National Power plc* [2000] IRLR 43.

Certain legislative changes might assist the courts and employment tribunals in arriving at a more effective approach to construction. In the context of the employer-employee relationship, the effect of the written statement law in section 1 of the Employment Rights Act 1996 is that documentation issued by the employer is not regarded as a definitive account of the contract; it is only the employer's view of the contract terms, and can be contradicted by other potential sources of the contract including collective agreements and custom and practice.[46] Thus section 1 is the functional equivalent, in employment law, of the principle, in consumer law, of judicial control of standard form agreements. Section 1, however, has no effect if the document in question is interpreted as giving rise to a relationship of self-employment. It makes no sense to allow the courts to intervene once the nature of the employment has been established, but to give the employer a more or less free rein to determine the legal status of the relationship in the first place. Thus there is a clear case for extending the courts' power to review and regulate standard form employment agreements to include control over documents which purport, directly or indirectly, to determine employee status.

[46] *System Floors (UK) Ltd v Daniel* [1982] ICR 54; *Robertson v British Gas Corp* [1983] ICR 351; Deakin and Morris, *op cit*, at pp 256–258.

16

Super Priority for Asset Acquisition Financing in Secured Transactions Law: Formalism or Functionalism?

CATHERINE WALSH*

I. INTRODUCTION

SECURED CREDIT IS a well-known and deep-rooted phenomenon. The basic idea is an intuitively simple one. Security gives a creditor the right, should the debtor default, to satisfy its claim out of the value of those assets that the debtor previously agreed would be charged with security, in preference to the claims of the debtor's other creditors.

The institution of secured credit has existed as long as notions of private property and freedom of contract.[1] However, for a long time, the concept of real security remained tied to what a common law lawyer might term the 'mortgage on Blackacre' model. Although a secured creditor's interest is ultimately in the liquidated value of the charged asset, not the asset itself, security was conceived in traditional property terms as a real right held by a person in an identified object, not future assets, and not generic funds of circulating assets such as inventory and the receivables derived from its disposition.[2]

In order to accommodate the demands of commercial financing, the collateral specificity doctrine—as applied to movable assets—has undergone widespread erosion. Over the last century or so, most western legal systems have come to accept some variation on the idea that an effective security

* For very helpful comments on earlier versions, I am grateful in particular to Professors Lionel Smith, Roderick Macdonald and Ulrich Drobnig.

[1] Although security can also arise by operation of law in specified creditor-debtor relationships, this article is concerned only with consensual or conventional security, ie security created by private agreement between the debtor and creditor.

[2] The doctrine of specificity has particularly dominated (and to some extent still dominates) the evolution of secured financing law in France (along with Italy, Greece, and to a lesser extent the Latin American countries). Reform has been cautious, incremental, and more often initiated in response to foreign than domestic commercial pressures.

can be granted through a single juridical act in all of a commercial debtor's present and after acquired assets, or in generic categories of present and future assets (eg, all inventory, or all claims), or both. In some legal systems, the institution of sale has been pressed in aid to achieve this result, the best example being the fiduciary transfer of title and the fiduciary assignment of claims recognised by German law.[3] In others it has been accomplished through the emergence of a dedicated security device. The best examples here are probably the judicially sanctioned equitable charge of English law in both its fixed and floating manifestations,[4] and the statutory 'security interest' concept which provides the organising structure for Article 9 of the Uniform Commercial Code in the United States and the Canadian (and New Zealand) Personal Property Security Acts. [5]

Despite the emergence of what will be termed a generic concept of real security, security continues to be conceived of as a real right—the property element is essential to maintain the priority and enforcement advantages traditionally associated with security.[6] Thus, security taken in a debtor's after acquired assets is still caught by conventional property theory to the extent that it cannot vest until the debtor actually acquires title to specific assets falling within the generic categories of after acquired assets over which security was granted.[7] But once this happens, the security, as a general rule, attaches with effect either from the date of the security agreement, or from

[3] The German developments are almost entirely a product of contractual innovation sanctioned by the judiciary in the absence of express codal provision for an effective form of non-possessory security in commercial assets. See eg, Jens Hausman, 'The Value of Public-Notice Filing under Uniform Commercial Code Art 9: A Comparison with the German Legal System of Securities in Personal Property' (1996) 25 *Ga J Int'l & Comp L* 427; Hein Kötz, 'Rights of Third Parties: Third Party Beneficiaries and Assignment' ch 13, v VII, *International Encyclopedia of Comparative Law* (J C B Mohr (Paul Siebeck), Tübingen, and Martinus Nijhoff, Dordrecht, Boston, 1990) at 98–99. Nonetheless, these forms of financing are widely used and recognized to be legitimate. They have been picked up by a number of other civil law countries, including Japan, Korea, Indonesia, Luxembourg, the Netherlands (superseded by the concept of the non-possessory registered pledge in the 1992 Dutch Civil Code) and Quebec (preserved in somewhat different form in the 1994 Civil Code of Quebec).

[4] For a concise exposition, see Lionel Smith, Security, ch 5 (pp 385–464), in Peter Birks, (ed), *English Private Law* (Oxford: Oxford University Press). For a comprehensive analysis, see Sarah Worthington, *Proprietary Interests in Commercial Transactions* (Oxford: Clarendon, 1996).

[5] UCC § 9–204 (confirming that a security agreement may create or provide for a security interest in after acquired collateral). Note that all references are to revised Art 9 which became effective 1 July, 2001. For the PPSAs see eg NB PPSA, s 13.

[6] Lionel D Smith, *The Law of Tracing* (Oxford: Clarendon, 1997) at 314.

[7] Thus, a security interest does not *attach* under the Canadian PPSAs and Art 9 until the debtor acquires rights in the collateral; see eg, NB PPSA, s 12(1)(b); UCC § 9–203(b)(2). The idea of attachment in the case of the English floating charge requires special mention. It is sometimes said that the charge does not attach, or does not attach to any specific assets, until it is converted into a fixed charge by crystallisation. According to Sarah Worthington, however, the chargee's proprietary interest is the same whether the charge is fixed or floating; the only difference is that with a floating charge, there is a licence in the debtor to deal with the assets until crystallisation: Sarah Worthington, *Proprietary Interests in Commercial Transactions* (1996) 82–100.

the date of its publicity (where this is required),[8] or some variant on these two events. It is this element of retroactivity that makes the grant of security through a single juridical act effective. Otherwise a fresh grant of security would be needed for each asset subsequently acquired by the debtor, with effectiveness against third parties only from the date of acquisition.

This chapter addresses an issue that confronts policy makers in any regime that recognises a generic concept of security. When and to what extent should a subsequent secured creditor who finances the debtor's acquisition of assets enjoy super priority with respect to those assets as against the holder of a prior generic security interest that seeks to charge after acquired assets of the same genre?

Although virtually every legal system recognises some form of super priority, there is wide divergence in its scope and extent. In some legal systems, super priority is available only to supplier-creditors—conditional sellers, hirers, financing lessors—through the device of retaining title to the supplied asset as security for the unpaid acquisition price. In others, super-priority benefits all asset acquisition financers so that a supplier of loan credit also prevails. In some, super priority arises regardless of the nature of the collateral, whereas others distinguish the pre-conditions for super priority depending on whether the collateral is an item of capital equipment or inventory. Finally, in some, super priority is limited to the acquired assets, whereas in others it extends to the proceeds of their disposition by the debtor, an issue of particular practical significance in the case of inventory acquisition financing.[9]

At a time of unprecedented interest in the modernisation and harmonisation of the law of real security, some level of policy convergence on such a basic and ubiquitous issue of commercial law would be welcome. This chapter seeks to contribute to that goal by analysing the possible reasons for the divergent approaches reflected in the comparative data.

In the interests of economy of detail, Canadian secured transactions laws will be used to provide the principal comparative data. Over the last three decades, each of the ten Canadian provinces and the three territories has undertaken a fundamental 'modernization' and 'rationalization' of the law of secured transactions. In the nine common law provinces and the three federal territories, the reform product is captured in a statute—uniformly entitled the *Personal Property Security Act* (PPSAs)[10]—which finds its intellectual roots

[8] See eg, NB PPSA s 35(1) (priority among non-possessory security interests governed by the order of registration; priority between a possessory and a non-possessory interest governed by the order of taking possession and registration).

[9] The extension of the acquisition financer's security over raw materials into the products of their manufacture or processing raises similar issues. However, in the interests of manageability, the chapter deals only with proceeds.

[10] In order of implementation, see: Ontario, 1976 (SO 1967, c 73, in force 1 April 1976, replaced by SO 1989, c 16, in force 10 October 1989); Manitoba, 1978 (SM 1973, c 5, in force 1 September 1978, see now R S M 1987, c P35); Saskatchewan, 1981 (SS 1979–80, c P-6.1, in

in the English floating charge and Article 9 of the UCC. In the civil law province of Quebec, the reform occurred as one element of a comprehensive overhaul of the whole of private law, and one therefore looks to the provisions on hypothecs and title-based security devices contained in the 1994 Civil Code of Quebec (CCQ).[11]

The Canadian reforms were implemented incrementally on a jurisdiction by jurisdiction basis over some three decades. One might suppose therefore that an implicit process of intergovernmental competition would ultimately have yielded a consensus on the best rules. One would be wrong. The treatment of acquisition financing super priority varies not just as between the CCQ and the PPSAs but even (though to a much lesser extent) among different versions of the PPSAs.

The Canadian experience is humbling for those who might reasonably expect policy convergence in relation to a kind of transaction—basic acquisition financing—which has essentially the same economic function in all markets and is therefore least likely to be affected by purely localized factors in theory. What I want to explore in the balance of the chapter, are the implications of the Canadian experience. What has prevented convergence in relation to such a fundamental area of commercial practice in what is basically an integrated national credit financing market? Or is there an underlying *de facto* convergence despite the apparent rule diversity?

II. ASSIMILATION OF SUPPLIER AND LENDER SUPER PRIORITY

The juridical origin of super priority for acquisition financers everywhere is rooted in the idea of retention of title by suppliers as security for the unpaid acquisition price pursuant to a conditional sale, hire purchase, financing lease or similar transaction. The rationale for super priority in this context was—and continues to be in many systems—expressed in the internal logic of property law. The supplier's retention of title means that the asset never enters the debtor's patrimony, or does so subject to the seller's prior ownership rights, and thereby avoids the grasp of a prior generic security interest covering the debtor's after acquired assets.

force 1 May 1981, replaced by SS 1993, c P-6.2, in force 1 April 1995); Yukon Territory (OYT 1980, c 20, 2d Sess, in force 1 June 1982, see now RSY 1986, c 130); Alberta (SA 1988, c P–4.05, in force 1 October 1990); British Columbia (SBC 1989, c 36, in force 1 October 1990); New Brunswick, 1995 (SNB 1993, c P–7.1, in force 18 April 1995); Nova Scotia, 1997 (SNS 1995–96, c 13, in force 3 November 1997); Prince Edward Island, 1998 (SPEI 1997, c 33, in force 27 April 1998); Newfoundland, 1999 (SN 1998, c P–7.1, in force 13 December 1999); Northwest Territories (SNWT 1994, c 8, in force 7 May 2001); Nunavut (Nunavut Consolidated Acts, in force 7 May 2001).

[11] SQ 1991, c 64, in force 1 January 1994. See in particular: arts 2660–2802 (hypothecs); arts 1745–1749 (installment sales); arts 1750–1756 (sale with a right of redemption); arts 1842–1850 (leasing-credit-bail); arts 1851–1853 (lease-*louage*); arts 2934–3006 (publication of rights).

In legal systems that have embraced the kind of functionally oriented legislative scheme associated with UCC Article 9 and the Canadian PPSAs, security is conceived in broad terms as encompassing all transactions that function to secure debt regardless of whether title vests in the debtor or the secured creditor.[12] Under this approach, the same legal framework that applies to conventional security constituted by a grant from the debtor is extended to security constituted by a reservation of title by the creditor on the theory that both serve the same economic function; that of securing the repayment of debt through the vesting of a property right in the creditor.

I have (in company with others) previously questioned whether the assimilation of title retention security to conventional security is as functionally justified as most article 9 and PPSA analysts assume.[13] Nonetheless, for the purposes of this chapter, I am content to assume that because buyers and lessees under a retention of title financing agreement acquire most of the indicia of beneficial ownership from the outset of the transaction, they should be afforded the same protection as regards their default rights and remedies as debtors under a conventional security agreement. I am similarly content to assume that because retention of title type security devices raise the same publicity concerns for third parties as conventional forms of non-possessory security, they should be subjected to the same registration or other publicity requirements that condition the third party effectiveness of conventional security.

However, issues of publicity and enforcement are pervasive issues of secured transaction law not unique to asset acquisition financing. What I am concerned with in this chapter is a second level of functionalism at work in Article 9 and the PPSAs. This is the assimilation of conventional security to retention of title devices for the purposes of extending the super priority traditionally enjoyed by suppliers who reserve title to conventional lenders to the extent the loan is intended to, and in fact is applied, to finance the acquisition of the collateral in which security is granted.[14]

The assimilation of vendor and lenders for the purposes of acquisition financing super priority is typically seen as just another variation of the functional premise associated with the decision of the Article 9 and PPSA

[12] For the PPSAs, see eg NB PSSA s 1 (security interest), s 3(1) (irrelevance of location of title as between creditor and debtor to applicability of Act). For the UCC, see § 1–201, § 9–109(a) § 9–202.

[13] Michael G Bridge, Roderick Macdonald, Ralph Simmonds, and Catherine Walsh, 'Formalism, Functionalism and Understanding the Law of Secured Transactions' (1999) 44 *McGill LJ* 567.

[14] The PPSAs define a purchase money security interest (PMSI) as a security interest which secures payment of part or all of the purchase price of the collateral or a security interest taken by a person who gives value for the purpose of enabling the debtor to acquire rights in the collateral. See eg NB PPSA s 1 ('purchase money security interest'). A PMSI takes priority over a prior security interest covering after-acquired assets that include assets of the same kind as those caught by the PMSI: see eg NB PPSA s 34. For the UCC, see § 9–103, § 9–324.

drafters to regulate all interests that operate to secure debt identically.[15] Since the economic function and effect of acquisition financing is identical regardless of whether formal title to the collateral is located in the creditor or debtor, why should lenders not enjoy the super priority traditionally available to retention of title suppliers?

'Treat like cases alike' is a good enough justification so long as the two cases are indeed alike. However, unlike the assimilation of retention of title to conventional security for the purposes of publicity and enforcement, the assimilation is operating in the contrary direction when it comes to the super priority issue. Here, secured transactions law is seeking to assimilate a theory of super priority associated with sales law and derived from conventional ideas about the specificity of property and the preservation of ownership rights.

To re-emphasize what has already been emphasised, the rationale for super priority in the case of a conditional sale, hire purchase, financing lease or similar transaction in which title is retained as security for the acquisition price is derived from the logic of property ownership. A prior generic security interest can only capture after acquired assets that come within the debtor's patrimony. The supplier's retention of ownership keeps the asset (or, at any rate, the supplier's interest in the asset) outside the debtor's patrimony. No additional explanation is needed. Thus, the drafters of the CCQ assumed that sellers and lessors who retain title to financed assets did not need explicit super priority protection against a prior hypothecary creditor holding security over the debtor's after acquired movable property; they prevail automatically by virtue of their status as owners.[16]

The 'treat like cases alike' principle justifies the export of a theory of super priority inherent in the logic of retention of title to conventional security if the secured creditor is a seller or lessor of the very asset in which security is taken. Otherwise, the supplier who happened to contract for the retention of a security interest as opposed to a retention of title would lose out to the prior holder of a generic security right covering after acquired assets of the same kind. In the relatively informal selling and leasing sectors, technicalities of this sort ought not to matter. Thus the CCQ awards explicit super priority to a vendor who takes a hypothec to secure the purchase price of the collateral[17] equivalent to that acquired automatically by a supplier who retains title as security for the acquisition price.

[15] See eg, CG van der Mewe and LD Smith, 'Financing the Purchase of Stock by The Transfer of Ownership as Security: A Simulated Transaction' (1999) *Stell LR* 303, 312.
[16] CCQ Arts 1745, 1750, 1842, 1851.
[17] CCQ Art 2954. However, the vendor hypothec enjoys super priority only if it the hypothec is created in the granter's act of acquisition, and published within fifteen days after the sale. This makes the vendor hypothec impracticable for most types of inventory financing because of the resulting need for successive registrations for each item of inventory. While the same problem would apply to inventory financing based on retention of title, the CCQ expressly

On the other hand, non-possessory hypothecary *lenders* under the CCQ are presumptively ranked according to the order of their registration of a notice of the hypothec.[18] The purpose of any later loans is irrelevant. The first registered hypothecary lender prevails as regards all after acquired assets that fall within the categorical boundaries of the classes of property caught by that hypothec. The fact that a hypothec that is later in time is granted to secure an asset acquisition loan does not accord it any special privilege.

According to the traditional logic of retention of title super priority, the CCQ's denial of super priority to asset acquisition secured lenders makes sense. Suppliers prevail over prior generic security interests by virtue of their retention of ownership. Lenders by definition are not owners and cannot rely on the super priority inherent in that status. On the contrary, for secured lenders, 'treat like cases alike' presumptively requires a temporal ranking.[19] If asset acquisition lending is to constitute some special exception to temporal ranking, the justification must come from the goals and policies of secured transactions law. Justifications derived from a supplier's retention of title simply have no application to lenders.

In the Article 9 and PPSA context, super priority for asset acquisition financing is typically defended as necessary to offset the situational monopoly on the debtor's credit needs that would otherwise be enjoyed by the first creditor on the scene to take and register security over the whole of the debtor's present and future movables assets, or significant chunks of them.[20] In the absence of an assurance of super priority, later lenders would be reluctant to advance financing. This is particularly true in systems in which the parties to a security agreement creating a generic security interest can agree that the collateral will cover all future lending between them.[21]

provides that a single publication suffices where the retention of title agreement is intended to cover after acquired inventory: Art 2961.1.

[18] CCQ Art 2945.
[19] Thomas H Jackson and Anthony T Kronman, 'Secured Financing and Priorities Among Creditors' (1979) 88 *Yale LJ* 1143, at 1144 ('The overriding priority afforded purchase money lenders occupies an especially important place in the Art 9 scheme because it constitutes an exception to the general rule that competing security interests are to be ranked temporally, with earlier interests prevailing over later ones.').
[20] Thomas H Jackson and Anthony T Kronman, 'Secured Financing and Priorities Among Creditors' (1979) 88 *Yale LJ* 1143, at 1171. The anti-monopoly justification is now commonplace. See eg, *18 Ariz J Int'l & Comp L 605* (2001), Meeting of OAS-CIDIP-VI, Drafting Committee on Secured Transactions, Model Inter-American Law on Secured Transactions and Commentaries, at 623–24: 'The main objective of the purchase money security interest (PMSI) is to provide freedom to the debtor to borrow funds from various lenders. The MILST creates a security interest that may cover a debtor's after-acquired property and proceeds. This type of security interest over future goods may tie a debtor to a particular lender and limit the debtor's ability to obtain credit from other sources. If this debtor is able to obtain additional financing for new goods, a PMSI provides the new lender with protection over previous secured parties with respect to those goods. In other words, a PMSI allows a party second-in-time to become first-in-right.'
[21] For the PPSAs, see eg, NB PPSA s 14; for the UCC, see § 9–204(c).

Even if the residual value of the collateral is substantial at the time of the application for subsequent financing, the later lender faces the risk that the first lender will extend further advances that will swallow up that residual value by the time enforcement becomes necessary.

As a self-sufficient justification, the anti-monopoly explanation is over-inclusive. Preserving debtor access to future sources of additional financing could be accomplished by any number of super priority rules in favour of any number of subsequent secured creditors. Why favour asset acquisition financers in particular?

Indeed, the anti monopoly justification is not just lacking in explanatory power; it is perverse. Awarding temporal priority to the holder of a generic security interest covering after acquired assets reduces priority risk and thereby reduces monitoring costs. This in turn benefits the debtor because the cost of credit will be correspondingly reduced (assuming a generally competitive credit market). Thus, by increasing priority risk, the inversion of temporal priority in favour of subsequent secured creditors presumptively undermines the benefit sought to be obtained by liberating secured lenders from the traditional mortgage on Blackacre model in the first instance.

To justify the inversion of priority for asset acquisition financers in particular, an appeal is sometimes made to ideas of unjust enrichment. It would be unfair, it is said, if a prior secured creditor holding security in after acquired assets were entitled to appropriate the value of an asset to which it has made no contribution.[22]

It is this kind of reasoning that underlies Anglo-Australian decisions suggesting that, as a matter of common law development, a super priority similar to that recognised by the PPSAs and Article 9 should be awarded to purchase money charges and mortgages as against a prior floating charge and even as against a prior fixed or prior crystallised floating charge.[23] This idea is captured in the following excerpt from an Australian decision:

> If this submission [that the prior all-assets charge should come ahead of the subsequent purchase money charge] were upheld, the commercial consequences could be quite staggering. Assume that a bank were to lend to company A the sum of $10 million. Assume that the company gave the bank a fixed and floating charge which was registered and which covered future assets. Assume the company borrowed $1 million from a financier in order to purchase a new plant. *On the [bank's] argument, equitable title to the new plant would pass to the bank even though it made absolutely no contribution to the cost of acquisition.* Any charge in favour of the actual provider of the

[22] See Steven Walt and Emily L Sherwin, 'Contribution Arguments in Commercial Law' 42 *Emory LJ* 897, for a critical analysis of the contribution justification for PMSI super priority.
[23] The leading cases in this line of authority are: *Re Connolly Bros Ltd (No 2)* [1912] 2 Ch 25; *Abbey National Building Society v Cann* [1991] 1 AC 56; *Sogelease Australia Pty Ltd v Boston Australia Ltd* (1991) 26 NSWLR 1.

purchase price would rank subsequently in priority to the charge in favour of the bank.[24] [Emphasis supplied.]

Like the antimonopoly justification, the contribution justification suffers from insufficient explanatory power. After all, any injection of new value by any subsequent creditor into a debtor's estate risks being appropriated to a prior secured creditor's generic after acquired property clause. There is no unfairness per se in this. The asset acquisition lender is as free as any other secured lender to refuse financing without an adequate guarantee of priority. In the absence of legal protection, an inter-creditor subordination agreement should be sought. Having failed to take this step, the acquisition lender implicitly accepts the risk of a loss of priority.

Indeed, the contribution justification is inadequate even if it were restricted to suppliers as opposed to lenders. After all, the PPSAs and Article 9 do not award any special super priority to unpaid suppliers who have not taken security even though they make an equivalent contribution to the debtor's asset base.[25] If the contribution justification were enough, the presence of security ought not to make any difference. After all, we are not attempting here to justify the priority of a supplier over unsecured creditors—where the additional property element associated with taking security is relevant—but rather the super priority of one class of secured creditors over prior secured creditors.

III. EQUIPMENT ACQUISITION SUPER PRIORITY

What is sometimes said to make the contribution principle as applied to asset acquisition financing uniquely defensible is that the first lender's loss of priority is limited to an asset which the debtor would not have acquired but for the acquisition lender's provision of financing.[26] Under this view, super priority depends on the assumption that otherwise the debtor would be

[24] *Sogelease Australia Pty Ltd v Boston Australia Ltd* (1991) 26 NSWLR 1.

[25] In view of the absence of any explicit statutory recognition of such an exception in Art 9 and the PPSAs, restitutionary arguments based on the new value contributed by unpaid unsecured vendors have met with no or only very limited success in eroding the priority enjoyed by a prior generic security interest covering after acquired assets of the same kind. For a comprehensive overview, see *Knox v Phoenix Leasing, Inc* 29 Cal App 4th 1357 (Cal Ct App 1994).

[26] See eg, Lionel Smith, Security, ch 5 (pp 385–464), in Peter Birks (ed), *English Private Law* (Oxford: Oxford University Press): 'Although in general terms a later security interest must be subordinate to an earlier one, in the case where the later lender is advancing 'purchase money' finance for an asset, there are good reasons for allowing priority in relation to that asset. This will remove what would otherwise be A's credit monopoly; at the same time it does not materially harm A, since the loss of priority is only in respect of an asset which would not have been acquired but for B's advance.' And see Henry Deeb Gabriel, Symposium: The Globalization of Secured Lending Law 'The New Zealand Personal Property Securities Act: A Comparison with the North American Model for Personal Property Security' (2000) 34 *Int'l Law* 1123, at

unable to finance its future asset acquisition needs. But this assumption holds true only if we further assume that the holder of a generic security over after acquired assets of the relevant kind is unwilling or unable to supply new financing to support the debtor's future asset acquisition needs. To the extent that this further assumption is correct, the denial of super priority would indeed be unfair to the debtor. Such a rule would amount to rewarding the proverbial 'dog in the manger' who refuses others what it does not want for itself.

Where the generic security interest covers acquired assets of the same kind for which future asset acquisition financing is needed, this indicates that its holder is at least presumptively willing to finance the future acquisition needs of its debtor. Applied to capital equipment financing, however, that presumption is open to reversal by the fact that in practice debtors often prefer leasing to the outright acquisition of ownership. The general lender is equipped to provide the latter type of financing but not the former. Indeed, it has a positive disinterest in taking on the additional burdens of ownership since this would amount to taking on the additional risks of equity financing.

If lease financing is to remain available for debtors, super priority must be retained for retention of title equipment financing conducted by lessors and indeed sellers in view of the often formalistic differences between sale and lease financing. However, where the source of the asset acquisition funding is a lender as opposed to a supplier, there is no basis for any similarly general assumption that if the law did not award super priority, the debtor would be unable to finance the acquisition of new assets. On the contrary, the very fact that the general secured creditor's security covers after acquired assets suggests that the general financing is presumptively meant to satisfy the debtor's commercial financing needs, including its future asset acquisition needs. To the extent that this is incorrect, why not cover this by an explicit subordination clause in the initial security agreement or by a subsequent inter-creditor agreement rather than awarding a general super-priority to all subsequent equipment lenders?

One possible answer lies in the fact that the need to preserve lease and sales financing reduces the credit cost advantage otherwise enjoyed by the holder of a generic security interest over subsequent asset acquisition lenders. The first in time secured lender will have to undertake additional background investigations in every instance where the debtor requests specific asset acquisition financing to determine whether the asset in

1126–29: 'Both PPSAs and Art 9 recognize that a security interest can be given on both property presently owned by the debtor as well as after-acquired property. While the secured lender's priority over the debtor's after-acquired property and proceeds enhances the position of secured creditors, it can seriously limit the ability of debtors to obtain financing of new assets unless the debtor is able to give to the new creditor a priority position superior to that of the holder of the prior security interest. Both the drafters of PPSAs and Art 9 recognize this, and the issue is resolved by giving the new financing creditors a 'super priority' over the prior secured creditor.'

question is being acquired by the debtor on a lease or sale basis. Otherwise, it runs the risk of providing acquisition financing in reliance on assets that the debtor is double financing through its suppliers by way of a retention of title credit structure.

Because the ability of the prior general secured creditor to rely on specific equipment in extending fresh credit would thus involve an increased investigatory burden that first in time priority is designed to alleviate, the cost to the debtor will be roughly the same whether the equipment financing is provided by the first in time lender or by a subsequent lender. This being the case, there is no reason to oppose extending super priority to equipment lenders in addition to sellers and lessors.

Moreover, the commercial reality is that a lender can acquire the super priority enjoyed by retention of title creditors through a variety of tripartite arrangements, for example, through an assignment of the supplier's rights to payment to the financing bank, or through a sale of the equipment to the lender who in turn retains title in a financing transaction with the debtor.[27] The fact that these mechanisms are prevalent in practice suggests that asset acquisition financing from lenders, if not strictly necessary, is at least highly desirable. Why force lenders into cumbersome and risky transactional structures when the direct grant of lender super priority would accommodate the financing needs of both suppliers and their customers at lower cost and with greater flexibility?

Super priority for both lenders and vendors offers the additional advantage of facilitating new sources of financing for debtors in financial distress. The debtor's existing general lender may be reluctant to take on additional credit risk, but without an assurance of super priority, a firm in financial difficulty will find itself less able to obtain financing from other sources for new equipment and is more likely to fail.[28]

IV. INVENTORY ACQUISITION SUPER PRIORITY

Most all-assets or after acquired property lending is based on a line of credit or similar revolving facility. To the extent the general lender relies on assets of the kind financed by a subsequent acquisition financer in adjusting the credit facility, super priority imposes an increased monitoring burden on the prior general lender that the grant of a first in time priority in after acquired assets otherwise alleviates. In order to guard against the risk of

[27] For a discussion of title based loan security devices, see generally CG van der Mewe and LD Smith, 'Financing the Purchase of Stock by The Transfer of Ownership as Security: A Simulated Transaction' (1999) *Stell LR* 303.

[28] Nicholas L Georgakopoulos, 'Bankruptcy Law for Productivity', 37 *Wake Forest Law Review* 51, Spring 2002. See also Steven L Schwarcz, 'The Easy Case for the Priority of Secured Claims in Bankruptcy' (1997) 47 *Duke LJ* 425.

extending new credit on the strength of assets that are in fact subject to an asset acquisition security interest, the general secured creditor will have to investigate the background source of all new assets.

The analysis to this point nonetheless supports super priority in respect of assets for which leasing may be an equally preferable financing structure: capital equipment. Although the grant of super priority to dedicated equipment financers means that the holder of a prior generic security cannot rely on this type of asset to support further advances, the limited scope of the inroad should not materially diminish debtor access to general secured financing.

The situation is very different in the case of inventory financing. There is no question that the holder of a generic security interest does rely, usually quite heavily, on inventory and the proceeds of inventory to secure its credit advances. Indeed, the collateral specificity doctrine was liberated in the first instance in order to accommodate the ability of lenders to provide a revolving credit facility against the security of such circulating funds of assets as a debtor's inventory and receivables. Thus, taken to its logical conclusion, the grant of super priority to subsequent inventory financers would effectively reverse the policy decision to move away from the mortgage on Blackacre model of real security in the interests of enhancing debtor access to secured credit at lower cost.

In fact the drafters of Article 9 and the PPSAs recognised the potential prejudice to prior general lenders created by the functional identity of generic security and dedicated inventory acquisition security. They sought to accommodate it by differentiating the conditions for PMSI super priority for inventory and non-inventory assets.

For non-inventory assets, the purchase money secured creditor has a grace period (fifteen days under the PPSAs, twenty under Article 9) to register (publicise) its security.[29] So long as registration occurs within this period, the PMSI holder prevails against the holder of a prior registered generic security interest.

For inventory, somewhat more onerous conditions apply.[30] Although the security can be publicised by a single registration covering both present and future inventory, there is no grace period for effecting publicity. Super priority dates only from the time of registration, and the inventory financer is additionally required to send an advance written notice to the holder of any prior-registered security interest covering after-acquired assets of the same kind, stating that it expects to acquire a purchase money security interest in inventory and describing the relevant inventory.

The advance notice requirement responds, and was designed to respond, to the increased risk for general secured lenders represented by purchase

[29] See eg, NB PPSA s 34(1). For the UCC, see 9–324(a).
[30] See eg, NB PPSA s 34(2). For the UCC, see 9–324(b).

money super priority in inventory in view of the functional identity of generic and inventory financers.[31] Advance notice ensures that the prior creditor is able to respond in a timely fashion to the increased threat to the value of its own collateral presented by the intervention of a new inventory financer, eg, by adjusting the level of subsequent advances, or calling a loan default if the potential risk is seen as too unmanageable and the subsequent inventory financer refuses to negotiate an acceptable subordination agreement.

The only other alternative would be for the first secured creditor to take a dedicated purchase money security interest of its own. However, dedicated purchase money financing imposes additional transaction costs on lenders. Unless restricted to specific brands of relatively high value assets, it is too burdensome for secured lenders to distinguish the application of the kind of line of credit financing they supply as between the debtor's inventory acquisition and other purposes. To impose an accounting obligation of this kind in order to maintain priority would effectively destroy the costs savings sought to be achieved by awarding temporal priority to the holder of a generic security over after acquired assets. The additional cost, for lenders, is worthwhile only in cases of debtor in distress financing.

On the other hand, why introduce the complications of super priority at all for inventory financers? In view of the functional identity of generic security interests and inventory financing security interests, why not simply revert to a temporal priority ordering? One possible answer lies in the desire to preserve supplier financing as an alternative to traditional bank credit in the interests not just of the debtor, but of the suppliers themselves for whom the ability to provide direct financing may sometimes be key to a sale. That Article 9 and the PPSAs recognise the rather more vulnerable position of suppliers as opposed to lenders is confirmed by the rules governing relative priority where both a supplier and a lender provide acquisition financing for the same asset. If the drafters were indifferent to the sources of financing, a pro rata sharing rule, or a first to register rule, would be the logical response. But priority is instead awarded to the supplier.[32]

[31] See eg, (2001) 18 *Ariz J Int'l & Comp* L 605, Meeting of OAS-CIDIP-VI, Drafting Committee on Secured Transactions, Model Inter-American Law on Secured Transactions and Commentaries, at 624 'The reason for this [advance notice] inventory-financing requirement stems from practices ... concerning lines of credit. US and Canadian practices allow borrowers with lines of credit to draw funds on the line of credit to the extent supported by an existing asset-base of inventory and accounts receivable. Under this arrangement, a debtor could increase its asset-base with inventory acquired from a supplier by using a PMSI arrangement. The secured debtor can claim to have a larger asset-base, by supporting this claim with inventory encumbered by the PMSI if the PMSI lender does not notify the previous secured party. In that instance, the debtor could draw disproportionately on the line of credit and deceive the first lender. Consequently, the MILST attempts to resolve this problem by excluding assets encumbered by a PMSI from the asset-base by requiring that a PMSI lender provide notice to the original secured party.'

[32] See eg, NB PPSA s 34(4). For the UCC, see 9–324(g).

V. EXTENSION OF SUPER PRIORITY INTO THE PROCEEDS OF INVENTORY

The distinction between inventory and non-inventory acquisition financing is helpful in analysing a final area of divergence among different regimes. This is the issue of whether an acquisition financer's super priority in the inventory should extend into the monetary claims (receivables) arising from the debtor's sale of the inventory in the ordinary course of business. A buyer of goods in the ordinary course of business normally takes free of a security right granted by the seller;[33] any other rule would destroy the marketability of the debtor's inventory. Thus, to maintain the value of its collateral, the inventory financer is dependent on having a substitute claim to the claims arising from sales of the inventory.

Under the PPSAs and Article 9, a security right in original collateral extends to its identifiable or traceable proceeds by operation of law.[34] The proceeds security right is accorded the same priority status as the security right in the original collateral.[35] It follows that the holder of a PMSI in inventory presumptively prevails over a prior registered security interest granted in after-acquired property of the same kind as the proceeds.

The CCQ differs from the PPSAs and Article 9 in neglecting to extend retention of title creditor super priority in inventory to the proceeds of its disposition. The inventory supplier can always contract for a separate security right in proceeds derived from the sale of the inventory. However, this will not normally provide super priority. Since the contest is no longer between an owner and a prior secured creditor, but between competing assignees or secured creditors, the supplier loses the super priority advantage inherent in its ownership status. In the absence of a positive exception, the supplier's priority will be tested by reference to the general temporal rules of priority applicable to competing assignees or secured creditors.

On the other hand, the CCQ does not require the supplier to also give advance notice to prior registered creditors holding hypothecary security in after-acquired inventory. As with equipment financed by retention of title, it is sufficient that the transaction be publicised by registration within fifteen days. The refusal to extend super priority into the proceeds greatly alleviates the potential prejudice that the absence of a positive notice might otherwise inflict on the holder of a prior generic hypothec covering inventory. In order to avoid erosion of the value of its collateral, the supplier will have to take steps to ensure that active control is exercised over the proceeds of the sale of inventory so as to prevent them from falling within the debtor's patrimony, eg, by requiring payment into a special bank account held in the name of the secured creditor. The need for this kind of control

[33] See eg, NB PPSA s 30(2). For the UCC see 9–320(a).
[34] See eg, NB PPSA ss 1 ('proceeds'), 28(1)(b). For the UCC see 9–102, 9–315(a)(2).
[35] See eg, NB PPSA s 34(1)–(2). For the UCC see 9–324(a)-(b).

to be exercised in turn operates to protect the prior hypothecary creditor who will likewise be unwilling to extend financing against new inventory without assurance that the proceeds of its disposition remain within the debtor's patrimony.

A similar policy informs the English and Australian case law on the effectiveness of a clause stipulating that a retention of title supplier shall be considered the owner or trust beneficiary of the proceeds of sales of the inventory received by the debtor. After a period of initial hesitation, the courts have held that the interest of the supplier in the proceeds is properly characterized as a charge, not ownership, regardless of the contractual language used, and is subject to the normal priority rules governing competing charges.[36]

Subject to compliance with any registration requirements applicable to company charges, the inventory supplier's super priority *qua* chargee in English and Australian law depends on whether the charge is characterized as fixed or floating.[37] If the debtor is permitted by the supplier to deal with the charged claims and the funds collected on the charged claims in the ordinary course of business, the charge is floating, and will be subordinated to the holder of a prior charge, whether fixed or floating. But if the supplier exercises factual control over how the debtor deals with the proceeds, the charge is fixed and takes priority over prior charges other than a prior fixed charge.

The control requirement evidences that the inventory financer is in fact relying on the proceeds for repayment of the acquisition price so as to justify super priority. Just as importantly, it removes the risk of detrimental reliance by the prior general secured creditor; the prior lender cannot rely to its detriment on claims over which the debtor has restricted power to deal.[38]

At first impression the PPSAs and Article 9 seem to create a far greater risk of prejudice for the prior holder of a generic security interest in so far as the inventory financer's super priority in proceeds vests automatically without regard to evidence of monitoring or control. However, the old floating and fixed charge distinction survives to protect the prior general secured creditor in so far as payments made to creditors, including payments on a line of credit facility, from proceeds collected on the sale of inventory are transmitted free of any security interest in the proceeds,

[36] Lionel Smith, Security, ch 5 (pp 385–464), in Peter Birks (ed), *English Private Law* (Oxford: Oxford University Press, 2000).
[37] Control as the key to the floating fixed charge distinction was confirmed in *Agnew v Commissioner of Inland Revenue* [2001] 3 WLR 454, PC.
[38] On the functionalism underlying floating charge priority theories, see Catherine Walsh, 'The Floating Charge is Dead, Long Live the Floating Charge—A Canadian Perspective on the Reform of Personal Property Securities Law' ch 3 in *Perspectives on Commercial Law*, A Mugasha (ed), (Sydney: Prospect, 1999).

including that of an inventory financer. [39] Thus in order to defeat the prior general financer, the inventory financer will have to monitor and control the debtor's dealings with the proceeds. In addition, under Article 9 and the western PPSAs, the acquisition financer will not be able to assert super priority in proceeds that take the form of accounts (or proceeds of accounts). In order to protect the reliance interest of a prior secured creditor with security in the debtor's after-acquired claims, the inventory financer's super priority does not extend to such claims.[40] The Atlantic Canadian and Ontario PPSAs do not contain any equivalent exception. However, the former require the inventory financer to give advance give notice of its intention to take security to a prior secured creditor holding security in the debtor's after acquired accounts.[41]

The German position on super priority in inventory proceeds seems at first impression to impose considerably greater risk on a prior lender holding a bulk assignment of claims from the debtor. Priority between competing assignees of the same claims is ordered temporally according to the time of each assignment. Thus the holder of a global assignment of the debtor's present and after acquired claims by way of security normally enjoys priority over all subsequent assignees. However, in cases where the subsequent assignee is an inventory supplier who holds an assignment of the receivables arising on the buyer/debtor's sales of the inventory, the German courts have held that the priorities should be inverted. This qualification to first in time priority is based on the theory that where a business makes a bulk assignment of future receivables to a bank in circumstances where both parties ought to know that the assignor's inventory suppliers will refuse to supply without an assignment of the claims arising on the sale of the inventory, the assignment constitutes a fraud on the suppliers and is invalid as unconscionable or against public policy.[42]

[39] See eg, NB PPSA s 31. See also *Flexi-Coil Ltd v Kindersley District Credit Union Ltd* (1993), 5 PPSAC (2d) 192, 107 DLR (4th) 129 (Sask CA) and see the subsequently amended s 31 of the Sask PPSA. For Ontario, see *Credit Suisse Canada v 1133 Yonge Street Holdings* (1998), 41 OR (3d) 632 (CA), varying (1996) 28 OR (3d) 670 (Gen Div); *General Motors Acceptance Corp of Canada Ltd v Bank of Nova Scotia* (1986) 55 OR (2d) 438 (CA).

[40] See eg, Sask PPSA s 34(6) which reads as follows: 'A non-proceeds security interest in accounts that is given for new value has priority over a purchase-money security interest in the accounts as proceeds of inventory if a financing statement relating to the security interest in the accounts is registered before the purchase money security interest is perfected or a financing statement relating to it is registered.' This approach applies to accounts in any form that are proceeds of inventory, including deposit accounts in a financial institution, except in Saskatchewan, which excludes deposit accounts from the application of the rule: see Sask PPSA s 34(7).

[41] See eg, NB PPSA s 34(2)(b) which provides that the notice to be given by a purchase money financer of inventory to establish super priority in both the inventory and proceeds must be given to; '... any other secured party who has registered, before the registration of the financing statement relating to the purchase money security interest in the inventory, a financing statement where the collateral description in the financing statement includes the same item or kind of collateral or includes accounts.'

[42] See Hein Kötz, 'Rights of Third Parties: Third Party Beneficiaries and Assignment' ch 13, v VII, *International Encyclopedia of Comparative Law* (J C B Mohr (Paul Siebeck), Tübingen, and Martinus Nijhoff, Dordrecht, Boston, 1990) at 98–99; Vincent Sagaert, 'Cour de

The German position is explicable when we consider that the priority of the first in time bank lender is not dependent on public registration. In the absence of public registration, a subsequent supplier of inventory will be unwilling to extend retention of title financing to buyers because it has no means of verifying whether and how many prior bank lenders may be in the picture who would take ahead of their own claim. Automatic super priority relieves the supplier of this worry. The absence of a public registry means that the impact on the prior generic lender is also less severe. Because of the risk of a prior secret fiduciary transfer of equivalent breadth, the generic security holder is already in a position where it cannot be absolutely assured that it is entitled to first in time priority, and must invest greater monitoring resources in any event.

Moreover, the priorities revert to their normal first in time ranking where these conditions do not exist—that is where access to subsequent supplier financing is not dependent on a right in the proceeds—or where the prior bank loan was made specifically for inventory financing purposes.[43] In other words, super priority is limited by the same concerns with protecting the reliance interest of a prior general lender whose ongoing financing is specifically related to and dependent on inventory and its proceeds that inspired the advance notice requirement for inventory PMSI creditors in the PPSA context and the denial of super priority in proceeds under the CCQ and in the English case law.

In other continental systems, notably France and Belgium, the retention of title supplier acquires an automatic legal right to proceeds under a theory of real subrogation: upon the sale of inventory subject to retention of title, the supplier's real rights are transferred to the proceeds so long as they remain identifiable as such in the debtor's patrimony. Since, under a theory of real subrogation, the proceeds are simply the substitute repository of the original inventory, the proprietary status of the supplier's real rights remains the same. Thus, the inventory supplier *owns* the claims and prevails in its capacity as owner over a prior general secured creditor or assignee.[44]

cassation francaise, 26 April 2000—priority conflict between the seller under title retention and the assignee of the resale claim' (2000) 6 *European Review of Private Law* 823, at 833. And see Jens Hausman, 'The Value of Public-Notice Filing under Uniform Commercial Code Art 9: A Comparison with the German Legal System of Securities in Personal Property' (1996) 25 *Ga J Int'l & Comp L* 427. The exception does not apply where these conditions do not exist or where the bank loan was made specifically for inventory financing purposes. The German courts have also ruled that the supplier/assignee may be subordinated in turn to either a prior or a subsequent assignee of the same receivables who purchases the receivables under a non-recourse bulk factoring agreement.

[43] See Hein Kötz, 'Rights of Third Parties: Third Party Beneficiaries and Assignment' ch 13, v VII, *International Encyclopedia of Comparative Law* (J C B Mohr (Paul Siebeck), Tübingen, and Martinus Nijhoff, Dordrecht, Boston, 1990) at 98–99.

[44] This is the position in France and Belgium: see Vincent Sagaert, 'Cour de cassation francaise, 26 April 2000—priority conflict between the seller under title retention and the assignee of the resale claim' (2000) 6 *European Review of Private Law* 823.

Application of the real subrogation theory per se does not justify the generous proceeds rights accorded to retention of title creditors. A likely explanation lies in the fact that French and Belgian law still retain the doctrine of specificity to some degree with the result that there is less potential for conflict between the reliance interests of the supplier and prior lenders. Moreover, the proceeds right is limited by the requirement for identifiability and, unlike the situation under the PPSAs and Article 9, is lost on payment of the proceeds into a mixed fund.[45] This limitation in effect means that the inventory financer to be assured of protection will need to take control of the proceeds in a fashion similar to that required by English law and CCQ to achieve super priority.

VI. CONCLUSION

In response to commercial financing needs, the concept of real security in movables has moved from the traditional asset specific mortgage model to a more generic concept in which a secured creditor is able to take a first ranking security over the entirety of a debtor's present and after acquired movables assets, or significant generic categories, notably inventory and receivables. This article has examined the justifications said to underpin one of the best known exceptions: the super priority extended to subsequent secured creditors who finance the debtor's acquisition of new assets.

For suppliers, retention of title is the traditional root of super priority. The supplier's continued ownership prevents the asset from falling within the embrace of a prior secured creditor's generic after acquired property clause. Retention of title by definition does not explain the decision of the Article 9 and PPSA drafters to extend super priority to asset acquisition *lenders*. Nor can it explain why super priority should extend to the *proceeds* of the debtor's disposition of inventory subject to an asset acquisition security interest, whether taken by a lender or a supplier. In fact, different legal systems take divergent positions on both these issues.

In an effort to understand the reasons for the existing rule diversity, this article began by examining the traditional justifications advanced for awarding super priority to lenders as well as suppliers. It was found that the familiar anti-monopoly and restitutionary explanations lack sufficient explanatory power and that a better understanding is achieved if we instead focus on the debtor's specialised asset acquisition needs and the relative efficiency of different credit sources to respond to those needs. This leads to a distinction between financing capital equipment and financing inventory.

[45] Vincent Sagaert, 'Cour de cassation francaise, 26 April 2000-priority conflict between the seller under title retention and the assignee of the resale claim' (2000) 6 *European Review of Private Law* 823.

For capital equipment, it was hypothesised that super priority constitutes a necessary exception because of the established preference of commercial debtors for flexible lease and equivalent asset specific financing techniques, and the relatively minor prejudicial impact that the inversion of priorities has on the debtor's access to general financing from a prior secured creditor who has undertaken to supply general ongoing financing to the debtor on the strength of a generic security interest over the debtor's after-acquired assets.

The situation is very different for inventory. The holder of a first in time generic security interest inevitably relies on inventory and its proceeds in adjusting the debtor's ongoing general financing needs. Indeed, the need to facilitate the debtor's ability to use its inventory and receivables as security is the very reason why the concept of a generic security interest over generic categories of after acquired assets emerged in the first instance. So, for inventory, a super priority rule for subsequent acquisition financers is presumptively illogical: it is the exception that threatens to swallow the rule.

In view of the close functional identity between a general security interest over after-acquired assets of a debtor, and dedicated inventory financing, it was therefore hypothesized that the emergence of a generic concept of security would lead to increased qualifications and restrictions on the special priority status of dedicated inventory financers, including the traditional retention of title inventory supplier. This expectation is borne out by the comparative data.

Fully fledged super priority for inventory financing is confined to legal systems that either have not fully embraced the generic security interest concept (France, Belgium) or have done so without providing any system for the public filing of generic security interests (Germany). In a system that retains collateral specificity, super priority for inventory suppliers who retain title does not involve the same potential for conflict with the collateral reliance expectations of prior secured creditors. In a system that recognizes the generic security concept but does not provide public filing, debtor access to future supplier financing for inventory is necessary because the supplier will not be willing to supply in the absence of a reliable means of verifying, and taking steps to protect itself against, the possible priority risk posed by the holder of a prior generic security interest.

In contrast, in systems that have introduced both generic security and a system of public filing, one finds that super priority for inventory asset acquisition is qualified in a manner that seeks to protect the ability of the prior generic secured creditor to rely on incoming inventory and the proceeds of its disposition in making ongoing adjustments to the debtor's financing demands.

In an article 9 and PPSA type regime, this is accomplished by requiring inventory financers to give advance notice to the holder of prior generic security interest of the intention to take a security interest in inventory. This permits super priority to be awarded to the subsequent inventory financer while also protecting the prior secured creditor against the risk of extending further financing against the new inventory in the belief that it is unencumbered.

Under the CCQ, and in English law, the solution involves restricting super priority to retention of title suppliers and then limiting their priority to the inventory itself, and not also the proceeds. Super priority as to proceeds is available only if the supplier closely monitors the debtor's use of the proceeds so as to effectively prevent them from coming within the debtor's control, thereby pre-empting the risk of detrimental reliance on the new value represented by those proceeds on the part of the holder of a prior generic security interest.

The article 9 and PPSA advance notice requirement appears to offer a more flexible and balanced solution in so far as it protects the reliance interest of the holder of a generic security interest while still giving the subsequent asset acquisition financer the potential for super priority in both inventory and proceeds. However, appearances are deceptive. Under article 9, super priority does not extend to the monetary claims owing on the sale of proceeds, so that the prior generic secured creditor can continue to rely on having a first priority in the claims derived from the debtor's inventory; the same policy is adopted by a significant number of the PPSAs. Super priority is effectively restricted to cash proceeds, and then only if these remain identifiable or traceable, forcing the inventory financer to monitor the proceeds closely to preserve its super priority. Thus, the net result is not much different from that which obtains under English law and the CCQ.

If there is any single prescriptive lesson that emerges from this analysis, it is this. The benefits in terms of greater access to credit achieved by facilitating a first in time priority for the holder of a generic security over the debtor's after-acquired inventory do not coexist easily with the facilitation of debtor access to subsequent dedicated inventory acquisition financing. Protection of the reliance interest of the first in time generic secured creditor requires the imposition of qualifications on the existence or scope of super priority enjoyed by subsequent inventory financers. For inventory suppliers, it follows that retention of ownership no longer constitutes the all powerful source of priority over competing claims that it once did. Once generic security is permitted, suppliers are made more and more likely to be subjected to the same rules that apply to conventional secured lenders even at the priority level. What is potentially troubling about this is that while generic security facilitates inventory financing, it does so only for conventional bank lenders; by increasing transaction costs, it has quite the opposite effect on access to inventory financing from suppliers. Generic security in effect gives a competitive advantage to financial institutions over suppliers. This is a good thing only if we assume that the conventional secured lending credit market possesses inherent efficiencies that are not capable of being matched by supplier financers, and that it is adequately competitive within itself to ensure a fair cost of credit for debtors. Jurisdictions in the process of reform need to think carefully about the extent to which their domestic credit markets possess these characteristics in determining the best way to proceed.

17

The Floating Charge—An Elegy

RIZWAAN JAMEEL MOKAL*

I. INTRODUCTION: THE MYSTERIOUS FLOATING CHARGE

WHETHER THEY PRAISE or (more frequently) condemn it, commentators do not generally realise how great a puzzle the floating charge is. Consider what we ordinarily expect of a security interest. Most obviously, the secured creditor would hope to enjoy *priority* in the proceeds of sale of the collateral over his debtor's other creditors. However, the floating charge does a very bad job, ranking behind not only a later fixed charge over the same assets, but also behind statutory preferential claims. Once the relevant provisions of the Enterprise Act 2002 come into force, the categories of preferential creditors will shrink.[1] However, a particular proportion of the value of the collateral subject to a floating charge will be 'ring-fenced' for general unsecured creditors,[2] leaving the floating charge holder's position no better off.[3] One might also wish a security interest to *encumber* the collateral, so that attempts by the debtor to pass on the assets subject to the charge would be unable to defeat the charge holder's rights. With a floating charge, however, the debtor 'may deal with or dispose of such property without the approval of, or even consultation with, the charge holder',[4] and as long as this is done in the normal course of business, the transferee's title is unburdened by the charge

*I am very grateful to Michael Bridge, Alison Clarke, Ian Fletcher, Look Chan Ho, Dan Prentice and especially Jay Westbrook and Sarah Worthington for discussions and comments on the chapter, and Sandra Frisby for helpful suggestions as to the historical relationship between the floating charge and receivership. The views expressed are mine alone, as are all errors.

[1] Enterprise Act, s 251.

[2] See s 176A of the Insolvency Act 1986, as inserted by s 252 of the Enterprise Act.

[3] See also Insolvency Act 1986, s 19(5), which grants the administrator the power to subordinate the claim secured by the floating charge to new monies advanced to finance the administration.

[4] V Finch, *Corporate Insolvency Law—Perspectives and Principles* (Cambridge, CUP, 2002) (hereafter, 'Finch') at 80. Finch's views will be taken as representative of the traditional understanding of the floating charge.

holder's rights. At the very least, a secured creditor would wish to be able to tell how much collateral it had been offered in order to carry out a *risk assessment* on the loan. However, the creation of subsequent fixed charges and the accumulation of new preferential claims can dilute the floating charge holder's security, as can the debtor's ability to alienate the collateral free of the charge. So the floating charge holder cannot even know which assets it has security over, and how much they are worth! No wonder this has been described as a 'key weakness' of this device.[5]

Security is also often said to bring efficiency benefits by allowing the secured creditor to monitor the assets subject to the charge, with the aim of deterring financial agency costs. The existence of security is said to facilitate this monitoring by creating 'focal points' for the monitoring effort: instead of having to keep an eye globally on the debtor's affairs, the creditor may focus simply on the presence, value, and use of the assets given as collateral.[6] This lowers monitoring costs, and on conventional reasoning, some part of these savings might be passed on to the debtor as lower interest rates. However, as Finch notes, floating charges are generally taken over the debtor's entire undertaking,[7] which means 'monitoring in order to detect misbehaviour or calculate risks could involve scrutinising the whole business'. In effect on this reasoning, much of the point of taking security as an aid in monitoring is negated. In the result, 'the floating charge may offer a relatively expensive method of securing finance'.[8]

So if the floating charge is such a thoroughly inefficacious way of securing a loan from the lenders' perspective, why is it more or less ubiquitous in corporate debentures?[9] Critics offer a worrying explanation. Not only is the floating charge a terrible security device, it is also said to be exploitative. It is 'a mechanism peculiarly conducive to the transfer of insolvency wealth from unsecured to secured creditors', and most likely 'designed' for this very purpose.[10] It is 'a charge upon all future property', sucking after-acquired assets into its ambit 'without the injection of fresh value by the original creditor'. This creates a 'windfall' for the charge holder and ensures that the rate on the secured loan is 'increasingly advantageous' to it.[11] These interest rates are 'excessively profitable'

[5] Finch at 103.
[6] See eg, S Levmore, 'Monitors and Freeriders in Corporate and Commercial Settings' (1982) 92 *Yale LJ* 49.
[7] Finch at 80.
[8] Finch at 92.
[9] See eg, Finch at 82 nn 100–1, *inter alia*, citing evidence that about 80 per cent of all debtor companies in one sample had granted floating charges.
[10] Finch at 102.
[11] Finch at 455; see also at 462. It has to be said this last point is quite suspect even on this line of reasoning: surely the interest rate would have been set in the first place in full anticipation of the fact that more assets might well be drawn in as collateral? It is therefore difficult to see how it could become '*increasingly* advantageous' to the creditor.

because the floating charge holder's ability to exploit its position ensures 'risks are loaded onto unsecured creditors'.[12] What is more, unsecured creditors may have insufficient notice of the effects of the charge, since it might be 'impossible to tell from the [company charges'] register how much the floating charges secure'.[13] This deceptiveness adds to the exploitative value of this device, which is sufficiently great to make it universally popular for powerful lenders (ie banks) despite its many disadvantages.[14]

This is a damning indictment. In response, I want to suggest that the floating charge is not malignant, merely misunderstood. The floating charge is unlike other security interests, first and foremost, in that it provides a very poor priority to its holder in the debtor's insolvency. Focussing on this and yet keeping in mind its ubiquity encourages a search for its true function. It is suggested here that a theory claiming to reveal this function could be considered successful only if it could explain the development of the 'lightweight' floating charge, the reason why floating charges are usually taken over the debtor's entire estate[15] and why they are often coupled with fixed ones, the distinction between fixed and floating charges, and the desirability or otherwise of restricting the priority position of the holders of the latter.[16] Removing the misunderstanding as to the actual role of the floating charge would also allow us to quiet concerns that its main value has been in its ability to exploit. However, while it might not have been exploitative in the past, it will be argued that its existence will become wholly exploitative after the coming into force of the relevant portions of the Enterprise Act. It will therefore be suggested that it should now be abolished.

II. PRIORITY AND THE FLOATING CHARGE

It would be useful to begin by pointing out that the indictment laid out above is self-contradictory. If criticism of the floating charge as to its efficacy in securing priority for its holder is correct, then criticism on the basis that it serves as a siphon for taking value away from unsecured creditors must be wrong.

It is submitted that not only is the floating charge generally ineffective in giving its holder priority over others, it is in fact not a priority-based device at all.[17] First of all, there is no doubt that the floating charge does not do

[12] Finch at 263.
[13] Finch at 456. This is not such a problem with fixed charges; see Finch at 459.
[14] See Finch's rather ambivalent conclusions at, eg, 102 and 104–5.
[15] Finch at 80 and 92.
[16] Such as that brought about by s 176A of the Insolvency Act 1986, as inserted by s 252 of the Enterprise Act.
[17] In fact, it was in an effort to convert it into a priority-based device that lenders invented the 'negative pledge clause'. However, negative pledge clauses are not registrable, so their registration does not fix subsequent lenders or purchasers of the collateral with constructive notice. So subsequent lenders or purchasers are not bound by such clauses unless they have actual notice of their existence.

much for its holder as far as priority is concerned. The well-known doctrinal facts about its ranking have already been rehearsed. What has not been done in the past is to view this ranking in the context of the empirical evidence available as to the recoveries of various classes of creditor in insolvency proceedings. Finch's recent discussion is a representative example. She faithfully reproduces the relevant statistics, which generally derive from the annual Surveys of its members by the Association of Business Recovery Professionals (ABRP),[18] and recently, from the study done by Julian Franks and Oren Sussman.[19] From the latter, Finch reports that 'recovery rates for banks [who generally take security] are 77 per cent compared with 'close to zero' for trade creditors and 27 per cent for preferential creditors'.[20] And according to the ABRP's most recent Survey, overall recoveries were 35 per cent for preferential creditors, 53 per cent for secured ones, and 7 per cent for unsecured ones.[21]

What Finch does not do, in common with most other commentators, is to draw out the implications of these figures for the debate about the nature of the floating charge. Important for our purposes is the recovery rate for preferential creditors, consistently revealed by these studies to hover around the thirty pence in the pound figure. We should recall that, under the 'absolute priority rule', creditors allocated a junior position in the statutory scheme for the distribution of insolvent estates are paid anything *only if* creditors senior to them have been fully paid off. Now it would be too quick to conclude from this that the average recovery on a floating charge, which ranks behind preferential debts, is zero. Unsecured creditors rank even further behind preferential creditors, yet their recovery is above zero at least in the ABRP study. (As noted, it is close to zero for firms in the Franks and Sussman data set.) However, a broader look at the pattern of distribution of recoveries overall in insolvency proceedings provides interesting insight. Preferential creditors get back everything they are owed in just over 25 per cent of insolvency proceedings, receive nothing at all in just under 25 per cent of cases, and recover less than five pence in the pound in the remainder.[22] So not only is there no recovery at all under a floating charge in almost 25 per cent of insolvencies (because preferential creditors receive nothing), there is virtually nothing paid because of the floating charge in another 50 per cent of cases (since preferential claimants get under 5 per cent). And what about the 25 per cent of insolvency proceedings in which preferential creditors receive their full amount, so that there

[18] See, eg, ABRP, *9th Survey of Business Recovery in the UK* (22 November 2001).
[19] *The Cycle of Corporate Distress, Rescue and Dissolution* (IFA Working Paper 306, 2000) (hereafter, '*Cycle*').
[20] Finch at 82 n 102.
[21] *9th Survey* at 18.
[22] *Ibid*, Chart 16.

would be a possibility of significant payments to floating charge holders as well? It is interesting to note that in these instances, the recovery rate for *un*secured creditors is high, at almost twenty-five pence in the pound. This compares not unfavourably with average recovery rates for both secured creditors—reported in this Survey to be about 53 per cent[23]—and certainly those for preferential ones. A likely explanation for this finding is that in such cases, a high return to preferential creditors indicates that the debtor's estate has sufficient value to pay back all or almost all of the debt secured by the *fixed* charge. Now since the floating charge would generally secure the *same* debt as the fixed one, having enough left over to pay off preferential creditors in full should be correlated with having sufficient value in the insolvent estate to meet a significant proportion of the unsecured debt as well (the debt secured by the floating charge already having been paid, entirely or almost entirely, under the fixed charge).

One immediate conclusion, then, is that the widespread image of the floating charge—as a device 'designed' to create 'windfalls' for its holder by voraciously 'transferring wealth from' and 'loading the risks of insolvency onto' unsecured creditors—is grossly distorted. From the very evidence relied upon by its critics, it appears the floating charge directs, at best, tiny amounts of insolvency value to its holders in three-quarters of all insolvencies, and does not transfer *any* wealth at all in another quarter of the cases. This also means the argument that this charge loads insolvency risk onto unsecured creditors finds little support in the facts. And since the floating charge seems to take little out of the insolvent estate in an overwhelming majority of cases, and given the conventional view that a loan secured even by a floating charge is likely to be at least somewhat cheaper than an unsecured one, there remains little basis for suggesting that the rates charged on loans secured by floating charges are excessive (at least on the grounds listed above). Another conclusion is that the fact that registrations of this type of charge do not reveal the amount secured by, or the particular assets subject to it from time to time, is not objectionable, since the practical impact of the charge—its ability actually to 'bite' on the collateral, to 'corner' it for its holder, to apply it towards the discharge of the amount secured—is insignificant in most cases.[24]

So whatever else the role of the floating charge may be, it is not to exploit unsecured creditors by siphoning insolvency value away from them by virtue of the *priority* of this device. In fact, as this discussion shows, priority is not a great virtue at all of this charge. This leads to another conclusion. Let us focus on a peculiar feature of the 'lightweight' floating charge, viz, that it is taken by a lender whose exposure is fully or at least

[23] *Ibid*. Needless to say, the difference between secured and unsecured claims should be expected to be reflected to *some* degree in the recoveries for these types of creditor.
[24] The role of registration is discussed further in the following section.

substantially secured by a fixed charge,[25] so that the floating charge is not required to provide it priority. In light of the quick discussion so far, it is suggested that lightweight charges are not some special category of floating charge. In fact, floating charges are *essentially* 'lightweight' in that they are not good at, and therefore not relied upon to, provide priority for their holders in the proceeds of sale of the collateral. It would require great faith in the *ir*rationality of lenders which are repeat players in the market, which lend to tens of thousands of companies a year, and which have accumulated expertise in dealing with the consequences of default,[26] to attribute to them the intention of taking floating charges in the hope of gaining priority. It is more credible to assume that they would have long learnt from experience that they would get back next to nothing on this type of security in at least 75 per cent of insolvencies, and would not need to get anything on it in most of the remaining.[27] In addition and crucially, they could not predict at the time of the loan agreement which of these groups a particular debtor would fall into, should it become insolvent. So it is not surprising that generally, floating charges are coupled with fixed ones.[28] The two perform different roles, and it is the *fixed* one which is relied upon to provide priority, encumbrance, and risk assessment. Note that on the priority-based view of the floating charge, it is difficult to provide a satisfactory explanation not only of the rationale for lightweight charges, but also of why floating charges are normally taken together with fixed ones.[29]

It is very important to be clear about precisely what is being argued here. The argument is *not* that having a floating charge *never* increases the holder's recoveries simply by virtue of the priority of this security. The evidence discussed above is by no means inconsistent with the proposition that in *some* insolvencies, the priority of the floating charge would cause a greater amount of value in the insolvent estate to be diverted to the holder

[25] See, eg, C Rumbelow, 'Mortgages by companies: A problem' [1989] *LSG* 32, and F Oditah, 'Lightweight floating charges' [1991] *JBL* 49.

[26] See, eg, the discussion of 'central rescue units' in *Cycle*, examined below.

[27] This is not a 'psychological' claim about the state of mind (whatever that means) of institutional lenders. It might well be, eg, that some bank managers etc. would state if asked that they included floating charges in security packages for the benefit of their employer because of the priority of this type of charge. That would be irrelevant to the present argument. The claim in the text here is the different one that the *practices* of institutional lenders (rather than *statements* on their behalf) reveal that they have taken on board the priority-independent nature of the floating charge. The practices relied on in support of this argument are precisely the ones discussed in the text, eg that floating charges are coupled with fixed ones, the sort of assets they are taken over, and that they are sometimes taken for purposes quite obviously unrelated to their priority. The claim is of course also that the analysis here coheres with the evidence available while the priority-based view of the charge does not, and also that it can—in a way that a priority-based view cannot—explain the structure of the relevant law both before and after the coming into force of the Enterprise Act 2002, eg about the provisions requiring registration and some of those governing administration; see below.

[28] See, eg, *Cycle* at 9 Table 4.

[29] Finch attempts to do so, eg, at 81–2.

of the charge than would be the case if the charge did not exist. What is in fact being argued is that this is both rare, and for that reason, also unpredictable. *At the time that the debenture was being issued*, it would be very difficult to determine whether this debtor—should it later have to go through a formal insolvency proceeding—would constitute one of those rare cases where the priority of the floating charge would inflate its holder's recoveries. Put differently, it would be difficult *ex ante* to predict whether the eventual insolvency of any particular debtor would constitute the case where the priority of the floating charge was both not *illusory* (because all of the debtor's property had been distributed to fixed charge holders and preferential creditors) and not *unnecessary* (because the debt secured by the floating charge had already been paid under the fixed charge). Floating charges may *result* in some recoveries in some minute proportion of insolvencies, but the dominant *reason* for their inclusion in a vast proportion of debentures that create security interests is not to ensure these rare and trivial recoveries. As explained, it would not be a rational strategy for a lender seeking to gain priority over other creditors in a useful way to rely on the floating charge.

III. THE LIMITED ROLE OF THE FLOATING CHARGE

So what *is* the floating charge for? It is suggested that it is a *residual management displacement device* which can work optimally *only* when it is part of a package of security interests covering the whole or substantially the whole of the debtor's property,[30] and generally, only when this package includes fixed security over assets strategically important to the debtor's business.[31] For reasons that will soon be apparent, someone holding such a security package will be referred to in this chapter as the debtor's 'main creditor'. We can start building up our understanding of all this by taking into account suggestions about lightweight charges, and others about floating charges in general made quite recently in the literature. First, in explaining the causes for the emergence of lightweight charges, commentators have seized upon the obvious and important connection between having a floating charge over substantially the entire estate of the debtor, and being able to block the appointment of an administrator to manage it. Despite having protected its priority through a fixed charge, the lender nevertheless values these powers, since they are crucial to controlling the timing and manner of realisation of the security (ie that under the fixed charge).[32]

[30] Compare Insolvency Act 1986, s 29(2). The argument here is that far from identifying some special category of floating charge, this provision describes the sort of security package that a floating charge must generally be part of, in order to be efficacious.
[31] A somewhat similar understanding seems to underpin comments by R Goode, *Principles of Corporate Insolvency Law* 2nd edn, (London: Sweet & Maxwell, 1997) at 204.
[32] Oditah, eg, provides an insightful discussion in 'Lightweight floating charges'.

Secondly and not just in the context of lightweight charges, two sets of commentators have recently highlighted a possible *control*-based (as opposed to priority-based) explanation of the floating charge, noting that at least some of its value might lie in being able to take control of the debtor by appointing a receiver. Armour and Frisby, in their illuminating discussion of receivership, have noted that it 'is conceptually possible to imagine a legal system in which a 'floating charge' offers its holder *only* control rights and no benefits in terms of priority'.[33] And in their empirical study of the cycle of financial distress in the UK's SME sector, Franks and Sussman point out that 'the crucial feature of the floating charge is that it grants the holder the right to take control of the firm in the event of default, through the appointment of an Administrative Receiver. Control of the firm allows discretion over whether to realize the assets by selling the firm as a going concern or liquidating it. These control rights can considerably influence the size of the proceeds accruing to the creditor'.[34]

This brings out an element central to an accurate understanding of the floating charge. This is the ability to replace the existing management with an outsider. Franks and Sussman focus on the benefits to the floating charge holder of the ability to do so, but there are obvious advantages for other 'stakeholders' as well. Suppose we divide the causes of default on part of the debtor broadly into two categories:[35] management-related (to do with the incumbents' 'irrationalities, lack of ability, failures of strategy and deficiencies of understanding', etc.),[36] and management-unrelated (external shocks, macroeconomic comparative disadvantages, inflation, over-regulation, delayed payments, government policies like those on taxation, etc.).[37] The former category is crucial. Surveys of insolvency professionals routinely identify management-related causes as being of the greatest importance in corporate failures.[38] So for example, one in two companies undergoing formal insolvency proceedings had suffered distress

[33] J Armour and S Frisby, 'Rethinking receivership' (2001) 21 *OJLS* 73 at 90. The thrust of the argument here is that such a legal system is not only conceptually possible, but in fact, given the generally poor and universally unreliable priority of the floating charge, is more or less our system.
[34] *Cycle* at 6 (footnote omitted).
[35] See generally Finch at 126–40.
[36] Finch at 259.
[37] *Ex hypothesi*, this latter category excludes 'external' factors which a reasonably competent management would have anticipated and taken steps effectively to deal with.
[38] Finch at 126 n 34, provides a useful summary. Perhaps this is to be expected: insolvency practitioners clearly have an incentive to reinforce the suggestion of their relative superiority over humdrum non 'professional' managers. This might encourage their services to be resorted to more frequently and at an earlier stage of the corporate distress cycle (and so perhaps for longer). Even if that is the case (on which no position is taken here), it is probably true that, for the reasons to be explained in the text below, management failings are the primary cause of a significant proportion (perhaps the majority) of corporate insolvencies.

in the past,[39] 'yet the company's directors still did not prevent insolvency'.[40] It is a reasonable assumption that at least some of these insolvencies could have been avoided by providing better leadership to the company. Indeed, insolvency practitioners claim that about one in four insolvencies could have been prevented if directors had sought timely advice, but since they did not, 'in four out of five cases there was nothing [that] could be done to save the *company* by the time an insolvency practitioner had been appointed'.[41] Subject to appropriate qualifications as to how much weight such findings should be given, we can postulate that *some* companies would be able to avoid insolvency through a change in management. In the case of such companies, other 'stakeholders' would generally be better off if such change could be brought about at the right time.

Now when a debtor appears to be on the verge of defaulting on its obligations, the preceding analysis provokes the question of how we might design a system which can distinguish between businesses that would benefit from a change of leadership (ie those beset by management-related problems) and those that would not, and once this information is at hand, how we might ensure that it could effectively be made the basis of a decision to act appropriately.[42] Gathering information is of course costly, and requires both accumulated expertise and influence over the debtor. Ideally, it would be based on (a) the debtor itself reporting potentially relevant information to an appropriate decision-maker, which in turn (b) has the competence to decide whether the information indicates management failings, and if so, (c) has the ability to remove the management without unnecessary expense and replace it with a better one. Let us examine each of these steps in turn.

First, then, consider the relevance of the fixed charge. Loans secured with a fixed charge are almost universally subject to covenants requiring that the value of the secured assets, or the income from them, or both, not fall below a certain multiple of the debt secured.[43] Especially when the charge is over strategically important assets, this creates the necessity for the debtor to explain to the charge holder the strategy pursued, business decisions made, and results obtained. Further, the fact that the charge encumbers the collateral means that attempts to dispose of these crucial assets would only succeed if the charge holder approves. This again necessitates the provision of information justifying the decision to sell.

[39] *9th Survey* at 2. Franks and Sussman report this figure to be just over one in three; *Cycle* at 7 Table 2.

[40] This comes (inevitably, since that is the main available source of information) from the ABRP's *9th Survey* at 2; note, however, that this seems to assume that managerial action *could* have prevented insolvency.

[41] *Ibid.* (emphasis added to emphasise that this should not be regarded as identical with the company's *business*, which might still be saved, as discussed below).

[42] This draws on Armour and Frisby, 'Rethinking receivership' at 82–6.

[43] D Citron, 'The Incidence of Accounting-Based Covenants in UK Public Debt Contracts: An Empirical Analysis' (1995) 25 *Accounting and Business Research* 139, especially at 144–5.

This minimises waste: remember that it is cheaper for the debtor to volunteer such information than for the creditor to have to extract it for itself. From the perspective of junior creditors, the benefit is that management's attempts to substitute assets (machinery for cash, say, or cash for employee-hours) so as to increase the riskiness of the debtor's activities—which would reduce the expected value of all their claims—become subject to external scrutiny. More relevant to this discussion is the fact that the flow of such information over a period of time would allow a picture to be built up of managerial competence, and this picture might be available for examination, should the debtor threaten to default on the secured loan. However, providing this information and obtaining the consent of the creditor to crucial decisions by persuading the latter of their wisdom, is costly for the debtor (not least in time and the marshalling of evidence and arguments). It follows that, especially for debtors in the SME sector, these costs would be justified only 'once over', ie for one 'main creditor' per debtor. Any duplication would generally be grossly wasteful.

This last point is reinforced by the fact that all this information provision and gathering is useful only if the chosen creditor has the ability to make use of it. Developing the competence to assimilate it and react appropriately would be costly, and economies of scale and the need for accumulated expertise ensure that only repeat players fit the bill. Even then, of course, the benefits of being able to tell badly managed businesses (whose value could therefore be increased by a change of management, to the advantage of all those with claims against it) from those troubled for other reasons, would have to be sufficiently large to justify this expenditure. A large number of dispersed creditors, each potentially entitled to a minute fraction of this 'surplus', simply would not have a sufficient incentive to invest in this type of monitoring. If this surplus is to be preserved, this point indicates the need for a repeat-playing creditor which provides a significant proportion of the debtor's credit needs and thus has a significant stake in any surplus from remedying the debtor's management deficiencies.[44] The most obvious such actors are banks, and as recent research highlights, there are indeed extensive monitoring arrangements within them. Monitoring takes place both at individual branches and at specialist 'central rescue units' operated by the banks. Firms sent to these units remain there for seven and a half months on average, and up to seventy-five percent of these are turned around. There is also strong evidence that banks often encourage firms to remove members of their management team, and further, that firms significantly improve their chances of being turned around if such changes are made.[45] There is little doubt, therefore, that

[44] This draws on Armour and Frisby, 'Rethinking Receivership' at 84–6, and the sources they cite at nn 84–9.

[45] This summarises some of the evidence in *Cycle*.

there is an investment on the part of main creditors both in information gathering at individual branches and central 'central rescue units', and in the development of expertise in distinguishing bad managers from good ones.

This leads to the third step in the problem mentioned above, that once information about management competence has been gathered and analysed, the party in possession of it should be able to take appropriate action on its basis. The potential problem here would arise if this creditor had to convince other creditors or a court of the desirability of replacing the existing management. This is a problem precisely for the reasons mentioned, that verifying the evidence in court might be expensive (at least in terms of time), and that other creditors might not have sufficient experience and expertise in being able to tell, say, a bad managerial team apart from a merely unfortunate one. Further, other creditors, who might only enjoy a tiny proportion of any surplus arising from the replacement of poor managers, would suffer incentives to hold out for side-payments from the main creditor.[46] Concentrating decision-making powers as to the future of the management in the very creditor which has built up and assimilated information concerning these factors would resolve these problems.

This gets us to the point where we must postulate the need for something like the institution of receivership.[47] But it does not explain the need for a floating charge. *Ex hypothesi*, the 'main' creditor would have *fixed* charges under which it would enjoy the right to appoint a receiver with respect to the property subject to those charges. However, this by itself might not be satisfactory. Recall from the discussion above that subjecting assets to fixed charges involves significant costs both for the debtor and the lender. The essence of the fixed charge is that its holder has *control* over the disposal or (through the operation of loan covenants) significant change of use of the collateral.[48] We have noted the benefits of this control, but it also significantly constrains the debtor's ability to use those assets in the normal course of business.[49] For some 'circulating' assets, such as certain types of machinery, trading stock, and receivables,[50] the cost of crippling the debtor by rendering it unable to proceed without the creditor's consent, would far outweigh the benefits of the latter's monitoring. So creating fixed charges

[46] Armour and Frisby, 'Rethinking Receivership' at 84 including n 74.

[47] See, eg, Armour and Frisby, 'Rethinking Receivership' at 86–91. The appointment of a receiver suspends the managerial powers of the debtor's board of directors; *Moss Steamship Company Ltd v Whinney* [1912] AC 254, 263 (though the case concerned a court appointed receiver); *Meigh v Wickenden* [1942] 2 KB 160, 166; and *Gomba Holdings UK Ltd v Homan* [1986] 1 WLR 1301.

[48] See, eg, *Agnew v Commissioners of Inland Revenue* [2001] UKPC 28, [2001] 3 WLR 454 at [32].

[49] This has long been recognised; see eg *In re Florence Land and Public Works Co* 10 Ch D 465, 541 *per* Sir George Jessel MR, and *Biggerstaff v Rowatt's Wharf Ltd* [1896] 2 Ch 93, 101 and 103, *per* Lindley and Lopes LJJ. Both cases are mentioned by Lord Millett in *Agnew* at [7].

[50] See, eg, Finch at 81–2.

over such assets would often not be a rational strategy. This creates a problem *ex post*, once there is both a default and good reason to believe the management should be displaced. The benefits of removing the existing management would only accrue if this removal applied to the *entirety* of the debtor's undertaking. This might be because of the synergetic values of retaining under one management precisely those assets which could not rationally be subjected to the draconian discipline of the fixed charge. And in any case, what would have been badly managed would be the entire business, including those assets; so the incumbent managers' powers would best be curtailed with respect to the entire business. This would benefit all those interested in the undertaking of badly managed companies by improving their chances of survival, or otherwise by ensuring the value-preserving disposal of their business.

This, it is submitted, explains why floating charges are ubiquitous despite their poor priority position: they represent the residual element in the set of rights required properly to displace a poorly performing management. This also explains why they are generally coupled with fixed charges: the latter provide priority, encumbrance, and risk assessment, and also ensure a supply of the information that is vital to the effective deployment of the floating charge. Further, this analysis reveals why floating charges are generally taken over the debtor's entire undertaking, or else used—in combination with a fixed charge—with respect to assets not subject to the fixed charge. Their role is to 'mop up' and divert away control over (and *not* priority in) assets which, while the existing management is doing well, would best be left at their disposal.

Finally, we can also now understand why the oft-repeated criticism that the deceptiveness of the floating charge has never appropriately been remedied because the mechanism chosen for this purpose—registration—has never been efficacious, is also misplaced.[51] In fact, registration is not meant to counter the imaginary deceptiveness of the floating charge. Registration serves the different purpose of revealing the identity of the main creditor of a company. Under the pre-Enterprise Act law, it would be important to know for this purpose whether someone has fixed and floating security over all or almost all of the company's property (rather than to know exactly how much this person is owed). This would allow co-ordination of efforts by creditors to monitor the company's managers in order to control financial agency costs and also to build up a picture of their general competence. One way in which this might happen would be for potential creditors—notably, a second bank that might otherwise have

[51] See, eg, Lord Millett in *Agnew* at [10]. His Lordship describes the requirement that floating charges be registered as of providing merely 'theoretical' rather than real benefit to trade creditors, the alleged victims of the floating charge's deceptiveness. For the reasons given in the text here, this is as it should be, since the perceived 'mischief' of the floating charge might also be merely 'theoretical', not real.

set up its own monitoring arrangements—to refuse to lend upon discovering the existence of a prior floating charge (or else to enter into some arrangement with the first bank so as to avoid wasteful duplication of monitoring etc.).[52] Now because of its poor and unreliable priority, a floating charge is unlikely to be taken in isolation, and would generally be coupled with other, fixed, security over some or other of the company's property. This is bolstered by statutory provisions discouraging the taking of floating charges unless the creditor offers new value.[53] A creditor seeking security is not going to offer new value without gaining priority, and it cannot gain priority in a reliable and useful manner with a floating charge. This further increases the probability that a floating charge would be coupled with fixed security. It follows that the declaration of a floating charge on the register is a good signal of the existence and identity of a creditor playing the role of main creditor.

Finally in this section, it would be helpful to consider the mechanisms whereby control is transferred away from the management of a troubled debtor.[54] The usual way of capitalising on the fruitful co-working of the fixed and floating charges in the past has been receivership.[55] The receiver may be appointed when there is a default on the secured loan. His appointment will accompany another significant event. Either the default itself, or his appointment, will crystallise the floating charge. Once in office, the receiver draws his powers from two sources.[56] His *in rem* powers 'are held in right of the debenture holder and derive from the security created by the debenture. They include the power to collect in the assets comprising the security, to possess, control and use those assets and to deal with and dispose of them, whether by way of sale, lease, charge or otherwise'.[57] His *personal* powers arise by virtue of the fact that he is regarded, by the terms of the debenture, as the debtor company's agent, and thus has the power on

[52] That one way of coordinating creditor actions is precisely to ensure only one creditor capable of monitoring the debtor lends to it is sometimes missed by commentators.
[53] Insolvency Act 1986, s 245, decreases the net expected utility of taking a floating charge without offering new value.
[54] This discussion draws on Goode, *Insolvency* at 215–7 and 237–9.
[55] For the purposes of this argument, no distinction need be drawn between administrative receivership, and the receiver and manager of the pre-Insolvency Act 1986 law, except to note that where there is some reference here to powers *presumptively* granted to the administrative receiver by the Act, it will be assumed unless otherwise stated that the receiver and manager would have had equivalent powers *explicitly* under the debenture itself.
[56] For a statement which comes close to eliminating the distinction between *in rem* and personal powers, see *Independent Pension Trustees Ltd v LAW Construction Company Ltd* 1997 SLT 1105, 1110, *per* Lord Hamilton. His Lordship stated that an appropriately drafted floating charge could in principle encompass all commercially significant rights and powers of the company vested in its board. This should be read subject to the qualification that, at the very least, the right to challenge the validity of the debenture itself must as a matter of conceptual necessity remain with the board.
[57] Goode, *Insolvency* at 237.

its behalf to carry on its business.[58] And in order to allow these personal powers to be exercised effectively, the law implies into the debenture a promise on the part of the company that its board would not interfere with the receiver's running of the company.[59] Importantly for our purposes, note the three main uses of the *in rem* powers derived from the *floating* charge by the receiver. First, and as already mentioned, it acts to divest the board of control over these circulating assets, since the crystallisation of the floating charge finally encumbers the collateral, and any subsequent unilateral attempts to pass those assets out of the ambit of the security would not succeed. Secondly, the crystallised charge takes priority over, say, charging orders. Were this not the case, unsecured creditors would have an incentive to rush to press their claims, thus causing the business to be dismantled and any going concern surplus to be lost. Because of the floating charge, however, the receiver is taken to be dealing with assets subject to the property rights of the charge holder, the party with the prior (because of its fixed security) and often primary interest in the collateral.[60] Thirdly, once the company goes into winding-up, the receiver's agency relationship with the company terminates. However, the receiver's *in rem* powers with respect to substantially the *whole* of the company's property persist (including over those assets which could not be subjected to a fixed charge), thus allowing him to continue deploying all of it in a value maximising manner. Having to hand some assets over to the liquidator might destroy synergies.

The *absence* of this feature would have several disadvantages in addition to those already noted. First, it would encourage other creditors to engage in rent-seeking behaviour by threatening to initiate a winding up, which would disrupt the receivership precisely by removing from the receiver's control assets not subject to the fixed charge. Secondly, and following from that, if the expected cost of making these side payments was greater than the expected benefit to the charge-holding 'main' creditor of investing in information gathering and assimilation (as a prelude to initiating receivership if necessary), it would create a disincentive for it to do so. To the extent that this diminished its ability to detect and control financial agency costs, this would be harmful to all the claimants as a group. And third, it might encourage the 'main' creditor to substitute liquidation for receivership as

[58] These powers include being able to raise or borrow money and grant security, appoint a solicitor, accountant or other professional, bring or defend actions, refer to arbitration any questions affecting it, effect and maintain insurance in respect of its business and property, use its seal, execute deeds or receipts on its behalf, appoint an agent, call up any uncalled capital, and change the situation of its registered offices, etc; see Insolvency Act, Sch 1.
[59] See, eg, *Gomba Holdings UK Ltd v Homan* [1986] 3 All ER 94 at 98.
[60] This point is sometimes missed. Note also that preferential creditors, who do hold claims enjoying priority to that of the floating charge holder, would generally have no incentive to act precipitously for the same reason, viz, that their claims have statutory protection over those of the charge holder. Given the unnecessarily perverse structure of receivership, however, this confidence is not always justified; see the discussion below of occasional opportunistic behaviour by receivers.

the chosen method of management displacement. This would be undesirable because it might be more expensive to prove to a court's satisfaction that the company's liabilities exceeded its assets or that it was unable to pay its debts as they become due.[61] Much more importantly, the ability to remove the management by relying on a 'technical' default on a covenant in the debenture (eg requiring the income from some assets not to fall below a particular level) might be lost. An important benefit of the floating charge is its contribution to displacing the incumbent management *before* the company becomes insolvent in a way verifiable in court (and therefore perhaps more irredeemably).

IV. AN OLD PROBLEM

The discussion so far enables a bit of new light to be thrown on an old problem: the troublesome matter of distinguishing between fixed and floating charges.[62] Confusion as to the whereabouts of the boundary between these is a consequence of the confused view that, in however weak a manner, the floating charge is about gaining its holder priority. The key to this problem once again is to remember that the floating charge is not a priority-based device. This is important because encumbrance is a function (among other things) of priority, encumbrance and control (of the encumbered asset by the chargee) are necessarily related, and control is precisely what distinguishes fixed from floating charges.[63] Let us see what this means.

Consider the relationship between two of the features usually associated with a security interest, that with respect to the assets charged, it grants priority to its holder over the debtor's other creditors, and that it encumbers the assets so that they cannot be put beyond the ambit of the security without the security holder's consent. Now it is not possible to conceive of an asset being encumbered without relying on the notion of priority. To say that a charge encumbers an asset is simply to say that if the asset is transferred without the chargee's consent, the chargee would continue to enjoy priority in any proceeds of sale,[64] and that in appropriate circumstances, the transferee's title to the collateral is subject to the chargee's interests. To encumber an asset *is* to 'entrench' the priority of the chosen claim with respect to the asset and its proceeds by giving the

[61] Insolvency Act, s 123. In many circumstances, this problem can be overcome by reliance on the effect of the s 123(1)(a) statutory demand. See also *Taylor's Industrial Flooring v M&H Plant Hire* [1990] BCC 44.
[62] Uncertainty on this point is bemoaned by, eg Finch at 103 and 305; see also the sources cited at n 154.
[63] As recently reconfirmed by the Privy Council in *Agnew*.
[64] See, eg, *Foskett v McKeown* [2000] 3 All ER 97 at 119.

claimant a veto over any attempt by the debtor unilaterally to disturb that priority. So understanding that the fixed charge is connected with priority—in a way that the floating charge is not—helps us distinguish between the two types of charge. A charge only encumbers collateral if it entrenches the priority of its holder's claim, and an asset can only be encumbered by the charge if the chargee is given control over the use and disposal of the asset.[65] Not to have control means not to have priority. The floating charge has neither one nor therefore the other.

Importantly, it also follows that there is a price the chargee must pay for securing priority with respect to the proceeds of the assets in the debtor's insolvency. In order to do so, it must create and operate the mechanisms necessary for monitoring the use and disposal of the charged assets, and this it would do by exercising control over the collateral.[66] This is a substantive requirement, not merely a formal one, so that if no control were in fact exercised by the chargee, then the charge would be characterised as a floating one regardless of the parties' professed intentions.[67] Having such control does not, of course, ensure that all fixed charge holders would effectively monitor their debtors. As mentioned above, monitoring in this context is about gathering information about the debtor's use of its assets, and about holding a veto over their disposal or (because of loan covenants) changes in the use made of them. But monitoring is also about being willing and able to assimilate this information and having the expertise to judge good management decisions from bad. The discussion above ties the creditor's acquisition of priority to its possession and exercise of the power to sanction disposal of the collateral or changes in its use. How wisely the creditor exercises these powers is a different matter. Be that as it may, we can see that granting priority does at least create the opportunity for the priority-holding creditor to monitor the debtor with a view to controlling financial agency costs, and perhaps also, to build up a picture of managerial competence.

This also shows why it should not generally be possible to 'separate' a book debt from its proceeds, in order purportedly to create a fixed charge on the former and a floating charge on the latter. We can understand these points through an examination of the notorious Court of Appeal

[65] Consistently with this, see Millett LJ in *Re Cosslett (Contractors) Ltd* [1998] Ch 495 at 510.
[66] This suggests one reason why 'negative pledge clauses' are not usually effective; these often seek to ensure priority for a creditor without setting up real opportunities for the creditor to monitor. The same hostility displayed by English law to priority without monitoring could perhaps be seen in its attitude towards retention of title clauses, especially those concerning materials not in the form supplied (since the chances of the supplier being able to monitor these effectively are particularly slim).
[67] See, eg, Lord Millett in *Agnew* at [27]: 'But the banks did not want to monitor the bank account and be required to give their consent whenever the company wished to make a withdrawal. They wanted the best of both worlds. They wanted to have a fixed charge on the book debts while allowing the company the same freedom to use the proceeds that it would have if the charge were a floating charge.'

decision in *Re New Bullas Trading Ltd*.[68] The debtor D had agreed to grant 'fixed' charges over all its book debts to creditor C, and to pay all proceeds into a designated account. By the time this money reached the specified account, the agreement creating the charges envisaged that one of two things would have happened. There would either have been an event such as an order for the debtor's winding up or a petition to put it into administration, or the levying of distress against its property by another claimant followed by a demand by C, in which case the money in the account would be subject to a fixed charge and would not be dealt with except as directed by C. Alternatively, there might have been no such event, in which case the money deposited in the account would be 'released from the fixed charge' and would become subject to a floating charge, allowing D to deal with it in the normal course of business. Speaking for the Court of Appeal, Nourse LJ held that:

> just as it is open to contracting parties to provide for a fixed charge on future book debts, so it is open to them to provide that they shall be subject to a fixed charge while they are uncollected and a floating charge on realisation.

The discussion above shows that this argument is a *non sequitur*. Taking the relevant issues in turn, three points should be noted. First, we should ask whether C was capable of effectively acting as the main creditor, able—in the way described above—to harness together fixed and floating charges in a mutually beneficial way. The answer must be in the negative. The charges granted to C were not the only ones at play. Prior fixed and floating charges had been granted to the very bank that operated the account into which the proceeds of book debts were to be deposited. The duplication inherent in granting C a similar package of security interests throws doubt on its usefulness. D was not a particularly large company, and so, as noted above, the investment required to derive utility from this sort of security package would only have been justified once over, ie, by one creditor. Any duplication of information provision and assimilation would have been wasteful for the parties concerned. This should lead us to expect that no such duplication would be present. What is more, the bank would have an inherent advantage over C in monitoring the deposits in and especially the withdrawals from the account that D held with it.

So, secondly and unsurprisingly, it is clear from the facts that C's security package was virtually devoid of monitoring potential. As discussed above, information about the use of assets *actually* subject to fixed security would be provided by the debtor when it sought the security holder's permission to dispose of those assets, or to exchange them for riskier ones. For book debts, this happens when they are converted into money in the debtor's hands, most obviously and importantly, by being collected. At this point,

[68] [1994] 1 BCLC 485.

the debtor comes in possession of cash that might be used in riskier or negative net present value ways. So it is here that outside monitoring would be most useful. The contested debenture in *Re New Bullas*, on the other hand, gave D the right to deal with the proceeds in the normal course of its business precisely at this point, thus obviating the need for it to provide any information to C. In fact, the agreement clearly envisaged that C would rely on monitoring done by the bank: it permitted the latter to furnish to C any financial statements and other information available to it about D's assets and liabilities. The fact that C had given no directions at all as to the book debts subject to the charge even when D became seriously distressed[69] supports the conclusion that (perhaps other than reliance on the bank's actions) C was not doing any monitoring of D. So third and finally, the Court's reasoning is fallacious. It is perfectly possible to create a fixed charge on future book debts by ensuring that, *inter alia*, the debtor accounts to the charge holder for their proceeds as and when they are collected. However, it is not possible for a charge to be 'fixed' if it does not ensure the provision of at least some information to its holder regarding use of the collateral.

V. THE CHANGEOVER FROM ADMINISTRATIVE RECEIVERSHIP TO THE 'NEW' ADMINISTRATION

It is important to note that the emphasis of the argument here is not on vesting control over the debtor in someone, like a receiver, who owes primary (or in many respects, exclusive) allegiance to the holder of the floating charge. None of the analysis above depends on any arrangement featuring such single-minded devotion to the latter's interests.[70] The focus, instead, is on *di*vesting the incumbent management of control. In fact, it is almost certainly the case that receivership, especially under the wholly malign influence of decisions like *Downsview Nominees*,[71] has developed features that, apart from being unnecessary to the mutually beneficial operation of the floating charge discussed above, are both exploitative and value-destroying.[72] All of these undesirable incidents of receivership have the same roots. They are manifestations of the motivation costs created by the fact that the receiver expects to benefit from repeat appointments sent his way

[69] See, eg, the account of the facts in the first instance judgment at [1993] BCLC 1389.
[70] Indeed, receivership may not necessarily have developed in a way which conflates the perfectly distinct functions of (a) managerial displacement, and (b) exclusive and unilateral pursuit of the charge holder's interests. So for example, the receiver is considered the agent of the company, and interference with his activities might cause the charge holder to become his principal and thus be liable for any default; see *In re Vimbos* [1900] 1 Ch 470.
[71] *Downsview Nominees Ltd v First City Corporation Ltd* [1993] AC 295.
[72] With respect and for reasons which follow, Armour and Frisby are at their least persuasive in what overall is a very persuasive paper when they argue against the point being made here; see the discussion in 'Rethinking Receivership' at 100.

by a small group of institutions which act as main creditors with respect to an overwhelming majority of companies undergoing formal insolvency proceedings,[73] who has predominant responsibility for the appointing creditor's interests, but the costs of whose actions—including irredeemably wasteful acts or omissions—is mostly borne by a separate group of claimants.

Receivership is harmful in at least three respects. First, there are the costs of the procedure itself, which amount *on average* to a quarter of the value of the assets in the insolvent estate.[74] A process that consumes such a large proportion of the value it is meant to be distributing to a pre-determined group of claimants is intrinsically absurd.[75] The wastefulness of the process is strongly indicated by the fact that when receiverships have been tendered out (thus focussing the burden of any wasteful behaviour primarily on the insolvency practitioner concerned), their costs have fallen dramatically.[76] Arguably, Parliament saw the potential for just this type of motivation cost to be acute in the institution of receivership when it provided a power for the liquidator to approach the court to set the receiver's remuneration.[77] However, the courts (here as in so much else to do with receivership) destroyed the usefulness of this provision by insisting that they would only intervene if the level of remuneration set in the debenture was plainly excessive.[78] Given that standard terms in debentures provide that the receiver's remuneration would be fixed by reference to the charging practices of the receiver's firm,[79] and given also that the motivation costs mentioned above could be expected to be endemic within the institution of receivership regardless of the receiver's firm, the practical effect of this approach was to render the statutory provisions nugatory. The result is that, even by itself, this externality associated with receivership is probably sufficient to eat up most of the efficiencies generated by the otherwise beneficial joint operation of fixed and floating charges.

Secondly, receivers are now known sometimes to engage in rent-seeking behaviour on behalf of their appointor by attributing some costs to the floating charge instead of the fixed one, with a view to inflating the recoveries under the latter.[80] This is a practice which does cause loss to be

[73] See, eg, D Milman and D Mond, *Security and Corporate Rescue* (Manchester, Hodgsons, 1999).
[74] Franks and Sussman, 'Resolving financial distress by way of a contract: An empirical study of small UK companies' (22 October, 2000), available at URL: www.ifk cfs.de/papers/franks.pdf at 32–3.
[75] By contrast, formal insolvency proceedings cost on average 13 per cent in Sweden and 14 per cent in the US; see *ibid*.
[76] *Cycle* at 15.
[77] Insolvency Act, s 36 (this provision makes no explicit reference to indemnity or reimbursement, though the argument in the text here applies to these as much as it does to the receiver's remuneration); see also Companies Act 1948, s 371(1).
[78] *Re Potters Oils Ltd* [1986] 1 WLR 210.
[79] Lingard, *Bank Security Documents* 3rd ed, (London, Butterworths, 1993), para 7–04.
[80] Franks and Sussman, 'Resolving' at 18.

moved from the party best placed to deal with it *ex ante*, to those worse placed to do so. This is doubly the case since it is unpredictable *when* such rent-seeking behaviour might benefit the main creditor and when in turn a receiver might resort to it. So the fact that it might happen is unlikely to bring any (even theoretical) compensating benefits, in the form of lower interest rates from the main creditor, say. Thirdly, and most generally, there is the receiver's astonishing privilege to inflict harm on those, other than his appointor, who are interested in the debtor's estate, even when this would bring no benefits to his appointor.[81] This simply cannot be justified, for reasons of simple consistency in addition to everything else. Consider the receiver's duty, owed to junior charge-holders etc., to obtain the proper market price for the charged assets.[82] Keeping aside historical reasons, there is only one justification for this duty. This is what we might call the principle against gratuitous harm: not to require the receiver to obtain a proper market price would be to allow him to inflict gratuitous harm to the interests of junior claimants, which by definition is without reason and thus without justification. But if this principle applies to matters concerning the sale price of the charged assets, there is simply no reason for it not to apply to other functions performed by the receiver. In any case, receivership, as governed by *Downsview* and others of its ilk, is probably the only legal institution where property subject to the rights of a group of individuals is placed under the control of a person *not* bound to them by a general principle against gratuitous harm. It is for this reason that decisions like *Medforth v Blake*[83] are so welcome.[84]

Be that as it may, taking all these factors together makes it clear that few have reason to mourn the passing of this value-destroying, exploitative and anomalous institution.[85] Importantly, however, we should keep in mind the clear distinction between the floating charge, and receivership, the latter being only *one* possible way in which the advantages of this type of charge might be captured. Nothing in this discussion detracts from the analysis above, which points to the benefit of a properly informed creditor being able to remove the under-performing management of a defaulting firm without having to expend resources persuading other creditors or a court of the wisdom of this decision.

I suggest that the recent legislative changes to administration have moved precisely in this direction.[86] Under the changes introduced by the Enterprise

[81] See, eg, Finch at 266–7.
[82] Confirmed even by *Downsview Nominees*.
[83] [1999] 3 All ER 97.
[84] Under the new legislation, the principle against gratuitous harm can now be found explicitly in Enterprise Act 2002, Sch 16 (hereafter, 'Sch 16'), para 3(4)(b).
[85] Enterprise Act 2002, s 250.
[86] So eg, Finch at 270–2, recognises (though not for the reasons suggested here) that the new administration procedure might be regarded as an improved and more 'inclusive' version of receivership.

Act 2002, an administrator might be appointed either by court order or out of court. The latter route is generally available only to the debtor company itself or its directors, and to the holder of a security package that includes at least one floating charge and which covers substantially the whole of the company's property, ie the main creditor. The court order route is open in addition to any of the company's creditors (and to the justices' chief executive for a magistrates' court). Consistently with the analysis above, three points are of interest here. First, the main creditor may continue to appoint an administrator on the basis of a 'technical' default under the debenture secured by the floating charge. It need not be demonstrated that the company is insolvent, whether the administrator is appointed out-of-court[87] or by court order.[88] As noted above, this would allow a poorly performing management to be displaced early enough for the debtor to have a better chance of survival. Contrast this with the fact that *anyone* else seeking to appoint an administrator through either route would have to rely upon the fact that the company is or is likely to become unable to pay its debts.[89]

Secondly, ordinary creditors would have to prove to the court's satisfaction that the administration order was reasonably likely to achieve the 'purpose of administration'.[90] Given the way the purpose of administration has been framed in the legislation,[91] this is likely to be a significant

[87] Sch 16, para 16.
[88] Sch 16, para 35(2).
[89] Enterprise Act 2002, Sch 16, paras 11(a) (court order) and 27(2)(a) (out-of-court appointment by the company or its directors). In the case of directors, this requirement probably exists to counter the motivation costs arising in a situation where the advantages to the directors of having a moratorium in place outweighed the costs of administration to *them*, even though administration was not in the interests of all the relevant parties as a group. The declaration of insolvency (or near insolvency) required of them serves to bring home to them the fact that the company would thenceforth be operated (at least until it recovered), not in *their* interests (say) *qua* shareholders, but in the interests of its creditors as a group.
[90] Sch 16, para 11(b) and the marginal note to para 3(1).
[91] The statutory objectives of administration have been defined and ranked in a very interesting way, and are worth producing here in their entirety. Sch 16, para 3 provides that:

(1) The administrator of a company must perform his functions with the objective of-
  (a) rescuing the company as a going concern, or
  (b) achieving a better result for the company's creditors as a whole than would be likely if the company were wound up (without first being in administration), or
  (c) realising property in order to make a distribution to one or more secured or preferential creditors.

(2) Subject to sub-paragraph (4), the administrator of a company must perform his functions in the interests of the company's creditors as a whole.

(3) The administrator must perform his functions with the objective specified in sub-paragraph (1)(a) unless he thinks either-
  (a) that it is not reasonably practicable to achieve that objective, or
  (b) that the objective specified in sub-paragraph (1)(b) would achieve a better result for the company's creditors as a whole.

burden for most creditors. Note that before ordering an administration, the court would have to be satisfied of one of three things. First, the petitioning creditor would have to show that the administration was required to rescue the company as a going concern, which in turn would require the demonstration that despite its insolvency, the company was still more valuable as a going concern and hence worth preserving,[92] *and* (at least by implication) that the directors were unwilling to or incapable of complying with their duty to direct their best efforts to bringing about a turnaround in its fortunes.[93] Merely showing that the company was unable to pay its debts probably would *not* do, since that would still leave open the possibility that its distress was due to management-unrelated factors which the current directors themselves could steer it out of.[94] Secondly, and alternatively, the creditor would need to convince the court that interposing administration before liquidation would achieve a better result for the company's creditors as a group, probably by preserving the company's business, in whole or part, as a going concern. Once again, this would obviously depend on demonstrating that the company's existing management was not complying with its legal obligations.[95] Competent managers could be relied upon to perceive and respond to the need either for a formal insolvency proceeding (liquidation or administration) or at the least, the benefit to the company of the moratorium obtained through, say, administration, on the pursuit of claims against it. Finally, and if neither of the above conditions was satisfied, the petitioning creditor would have to show that an administrator would be better able to realise the company's property in order to make a distribution to secured or preferential creditors as appropriate, than the incumbent management. For the reasons already given, this could only be done by showing that the managers were not taking every step they

(4) The administrator may perform his functions with the objective specified in sub-paragraph (1)(c) only if-
   (a) he thinks that it is not reasonably practicable to achieve either of the objectives specified in sub-paragraph (1)(a) and (b), and
   (b) he does not unnecessarily harm the interests of the creditors of the company as a whole.

[92] Otherwise, Sch 16, para 3(3)(b) would apply.

[93] The duty derives from, eg, *Kinsela v Russell Kinsela Property Ltd (in liq)* (1986) 4 NSWLR 722, 730, *per* Street CJ; quoted with approval by Dillon LJ in *West Mercia Safetywear Ltd (in liq) v Dodd* [1988] BCLC 250, 252–3. In some circumstances, s 214 of the Insolvency Act 1986 might also be relevant; see eg, the discussion in Mokal, 'An agency cost analysis of the wrongful trading provisions' [2000] CLJ 335, at 365–6. The statutory duty would not *generally* be relevant under this head of the 'purpose of administration' because administration here is meant to rescue the company, while s 214 applies only when there is no reasonable prospect of avoiding insolvent liquidation.

[94] This flows from the undeniable observation that replacing the directors with an administrator is costly, and that it would be unjustifiable to incur these costs if the incumbent management itself could steer the company out of crisis.

[95] This is where s 214 would certainly be relevant. We are now concerned with the situation where attempting a rescue no longer appears advisable (Sch 16, para 3(3)), or in other words, that there is no reasonable prospect of avoiding an insolvent liquidation. At this point,

reasonably ought to be taking to minimise loss to the company's creditors (including, if appropriate, the initiation of a formal insolvency proceeding).[96]

For most creditors, being able to muster the evidence and arguments necessary to meet this burden would be very difficult indeed, requiring at the least something equivalent to the Rule 2.2 report that had virtually become a *sine qua non* for a successful application for administration under the old law.[97] The central aims of obtaining something like this report are of course to gather information about the debtor, and about whether it would benefit from the change of management that administration brings about. If the arguments made above are correct, the requirement to have such a report prepared should be waived for the main creditor, since the latter could be presumed to have acquired precisely that information through the normal operation of his security interests. A report would be superfluous and thus wasteful both of time and resources. The new law is quite consistent with this analysis. Under it, the company's directors, and more significantly, its main creditor, are both allowed to proceed without a court order simply on the strength of a statement by the intended administrator that 'in his opinion the purpose of administration is reasonably likely to be achieved'.[98] This treats the directors and the main creditor as on par in being in possession of information as to whether the company's business would benefit from a change of management. Either can be relied upon to arm the intended administrator with this information, enabling the latter to form the opinion that the purpose of administration is likely to be satisfied. This is bolstered by provisions which make it an offence knowingly to rely on a statement in making an out-of-court appointment that is false and that the appointor does not reasonably believe to be true.[99] Crucially, of course, neither the company's directors nor its main creditor have to persuade the court that the purpose of administration is likely to be fulfilled. Instead, the burden of showing the nominated administrator's statement to this effect to be false lies on the party wishing to challenge it.

directors come under a duty to take every step they reasonably ought to, to minimise loss to their company's creditors. This here means, most obviously, attempting to preserve any going concern surplus when disposing of the company's business.

[96] Note that because of the reference to secured creditors here, there is only a partial overlap between satisfying this particular requirement, and showing a breach of s 214: the latter imposes no duty on directors with respect to secured creditors. See the discussion in Mokal, 'Wrongful trading', especially at 357–60 and 364–5.

[97] Insolvency Rules 1986, r 2.2. Note also Sch 16, para 36, which allows the main creditor to intervene in any application to appoint an administrator not made by it, and to ask for its nominee to be substituted in place of the one nominated in the application. Arguably, the former would be more likely to fulfil the purpose of administration than the latter, perhaps because of the superior information provided to him by the main creditor. There also seems to be some discouragement here for other creditors to invest (perhaps wastefully) in something like a Rule 2.2 report, since the person whose appointment this report would back might never assume office.

[98] Sch 16, paras 18(3)(b) (main creditor) and 29(3)(b) (directors).

[99] See Sch 16, paras 18(7) and 29(7).

So thirdly, and following from both of the points just made, there is a clear reliance on the main creditor to distinguish between a poor management, and a capable but unfortunate one. For the main creditor, the need to persuade the court that a particular debtor could benefit from a change of management is transformed into the need merely to demonstrate that the debtor is in default. The only hurdle that the main creditor must overcome in displacing its debtor's management by this route is exactly the one suggested here, viz, that there is a default on a secured loan, and that it considers itself able justifiably to back the statutory statement by the nominated administrator, viz, that to allow the management to remain in place would be to jeopardise either the company's survival or the interests of all, or if appropriate, some, of its creditors.

VI. WHY THE FLOATING CHARGE MUST NOW BE ABOLISHED

The argument so far has been 'positive' in the sense that I have attempted to show the role actually played by the floating charge under the pre-Enterprise Act law, rather than 'normative' in the sense of advocating some change in the law.[100] Finally, however, I want to go further and explain why this paper claims to be an elegy for the floating charge, and not merely a eulogy. I will do so through a three-pronged argument. It will be suggested first that the beneficial role played by the floating charge under the pre-Enterprise Act law has now been rendered useless. This is the demise announced in the title of this paper. Second, the Enterprise Act retains certain functions for the floating charge, but it will be suggested that these do not depend on anything to do with its essential nature. They could equally be performed in other ways, and so do not justify the retention of this device. And finally, it will be claimed that the unnecessary retention of the floating charge will now cause it to operate, if only in a very small proportion of insolvencies, in a wholly exploitative manner.

Taking these points in order, then, recall that before the new administration procedure was introduced by the Enterprise Act, the floating charge was performing two main functions with respect to distressed companies. When such a company defaulted on a secured loan and the main creditor formed the judgement that the default was due to management-related reasons, the (now crystallised) floating charge would be used to mop up control, proprietarily, over those assets which could not rationally be subjected to a fixed charge during the company's normal operation. Secondly, and should liquidation start once receivership was underway, the

[100] The 'scare quotes' in the text indicate my adherence to the view that invocations of the positive/normative distinction are often highly misleading. What we understand the law to be is inevitably influenced by our views of what it ought to be, though the two need not be identical.

floating charge would prevent the liquidator taking possession of these assets, thus preserving any synergetic value. The receiver's *in personam* powers to manage the business having ceased upon the initiation of winding-up, he would have to rely on the *in rem* powers gained under the floating charge. Note also that the main creditor would generally (though not always)[101] have preferred receivership, (a) because it could make an out-of-court (and thus non-wasteful) appointment of a receiver but not of an administrator, (b) because the receiver could apply the debtor's assets towards discharging the secured debt while the administrator could not, (c) because of the receiver's near-exclusive attention to the chargee's interests, and of the greater rent-seeking potential of receivership, and (d) perhaps because of the legal-commercial inertia which tends to favour familiar procedures over unfamiliar ones.

Contrast the present position. The new administration procedure creates a status which confers on its holder all the powers the floating charge would have conferred on the receiver, while removing the features that made the 'old' administration undesirable. The administrator acts as the company's agent,[102] his appointment divests the directors of control *globally* (and not, of course, simply over assets subject to fixed security),[103] and also prevents the appointment of a liquidator.[104] The administrator may now be appointed out of court without having to prove insolvency and even on the basis of a technical default, as noted, and he is now able to make distributions to creditors.[105] So all the mutually beneficial functions previously performed by the floating charge have now been rendered redundant.[106]

Second, however, the Enterprise Act retains two important roles for the floating charge. Most important is the way the Act continues to define the main creditor. For the purposes of the Act, this is someone with one or more debentures (a) at least one of which declares that its holder has the right to appoint an administrator, and (b) at least one of which contains at least one floating charge.[107] Now the first condition is unobjectionable,

[101] See, eg, *Cycle* at 8–9 including nn 3–4.
[102] Sch 16, para 69.
[103] Sch 16, para 67: 'The administrator of a company shall on his appointment take custody or control of all the property to which he thinks the company is entitled.'
[104] Sch 16, para 42.
[105] Sch 16, para 65; ordinary unsecured creditors may only be paid by court approval or if the administrator considers the payment is likely to assist in the achievement of the purpose of administration.
[106] What is absent from administration of course is most of the set of incentives for the office-holder to engage in rent-seeking on behalf of the main creditor, which, needless to say, is the very reason for the legislative rejection of receivership.
[107] Sch 16, para 14. Note that para 14(3)(b) seems to treat as the main creditor (and thus one with the ability unilaterally to displace the management) someone who *only* has floating charges over substantially the whole of the debtor's property but no fixed security. This is inconsistent with the analysis in this paper: in every realistic situation that comes to mind,

simply requiring a record of the creditor's claim to act as the main creditor for this company. However, the second condition makes less sense. We know from the analysis above that, if the purpose is to enhance the value of the debtor's business to all those with claims against it, a creditor should only have the unilateral right to displace the debtor's management if it had both the ability to gather and assimilate information about its competence, and the expertise to decide whether the removal of this management was warranted. There should also be some way of discouraging other creditors from engaging in opportunistic behaviour. Now the ability to gather information arises from the possession of *fixed* charges over assets strategically important to the debtor's business. The role of *floating* charges is very much a residual one, basically to ensure that the security package as a whole covers the debtor's entire estate and undertaking. This would enable the entire estate to be kept under one management, thus protecting any synergies. Recall also that as a rule, the purpose of a floating charge is *not* to secure priority for its holder, since its priority position is weak, generally unnecessary (because it most often secures the same debts as the fixed charges), and quite unreliable *ex ante*. It follows that since an alternative method has now been provided to divest the managers of control over essentially revolving assets, of preserving synergies, and of capturing the benefits previously associated only with an ensuing receivership, the floating charge no longer has any necessary role to play in our conception of the 'main creditor'.

The second role assigned to the floating charge by the Enterprise Act is of allocating priority as between those claiming to act as main creditors to a company. Priority of the floating charge for this purpose is by date of creation, but this may be altered by agreement among all those holding such charges.[108] Unless it agrees otherwise, the holder of any prior charge is given the right to a minimum of two business-days' notice if the holder of a junior charge intends to appoint an administrator,[109] and may if it so chooses ask the court for its own nominee to be substituted for the person appointed by the junior chargee.[110] Now it is entirely sensible for a mechanism to be provided which allows for the co-ordination of monitoring responsibilities, since, as noted, there should only be one creditor investing in the monitoring of most companies. A truly effective way of ensuring this

the floating charge could only properly work parasitically, as an appendage to some fixed security. However, this provision probably does little harm precisely for this reason: as explained above, creditors are highly unlikely to rely exclusively on floating charges for security. So if the analysis in this paper is correct, not too many administrators will be appointed under para 14(3)(b). The Act also relies on the floating charge in identifying the main creditor for the purposes of the substitution or removal of the administrator in Sch 16, paras 92, 94, 97, and 103.

[108] Sch 16, para 15.
[109] Sch 16, para 15(1).
[110] Sch 16, para 96. Note that the prior chargee seems entitled to this substitution as of right, since para 96 does not appear to be subject to any qualification. Contrast para 95.

would have been to grant unilateral management-displacement rights only to the prior-most chargee (identified chronologically as a default, and otherwise by agreement).[111] It could perhaps be argued that allocation of the right to determine the identity of the administrator (which is what the legislation is doing here) is an attempt to create a similar incentive structure. However, what matters is the chosen creditor's ability to gather information, and this is connected of course with the fixed charge, not the floating one. So the focus here in the legislation on the latter seems quite arbitrary, a hangover from the period when the floating charge was required to preserve the unity, upon management displacement, of the debtor's entire estate. The allocation of monitoring responsibilities and thus of the unilateral right to displace the incumbent management could just as easily be made by, say, attaching the declaration of the right to appoint an administrator[112] with the first fixed charge over some (or any of a specified list of) strategically important assets.[113] The floating charge simply is not required as a distinct entity, and the legislative focus on it is hard to justify.

Moving now to the third prong of the argument, it is suggested that this focus on the floating charge in the new law is not only useless but also harmful. Here is why. It has repeatedly been argued that the priority of the floating charge is weak, unreliable, and often unnecessary. It was also noted, however, that in some (on the basis of the evidence considered, minute) proportion of insolvencies, its priority position *would* cause its holder to receive more from its insolvent debtor's estate than it otherwise would have. And this is what makes the continued existence of the floating charge objectionable.

Remember that in the distribution of an insolvent estate, the priority of one creditor over a non-consenting other is a privilege. It must be *earned* in a way which brings benefits to *all* the relevant parties as a group. Let us call this the 'privilege view' of security interests. Traditionally minded English lawyers have attempted to provide other justifications that circumvent the privilege view, most notably those based on the freedom of contract of the chargor and chargee, and on the property rights of the former. For reasons well explored in the existing literature, neither holds water. Freedom of contract arguments 'have force only with respect to arrangements that do not create direct externalities. [When] the contract directly impinges on the rights of third parties, there is no prima facie presumption of freedom

[111] Note the presumption in favour of the first-in-time chargee: it would have been operating monitoring arrangements for the longest period, and could be assumed to have relatively the most complete knowledge about managerial competence. The presumption would therefore be that it was best placed to decide whether any default was due to management-related factors.
[112] Sch 16, para 14(2).
[113] The list might include things like 'any of the debtor's bank accounts' or 'machinery involved in the debtor's day to day business'. These suggestions are very tentative, and a lot of thought would obviously have to go into formulating something like this list.

of contract'.[114] The truth of this proposition and of its close variants is attested by diverse parts of the legal system, competition law being perhaps the most important example. The privilege view of (priority-based) security interests is an acknowledgement of the fact that they seem to constitute a violation of this principle. At least *prima facie*, the contract whereby A grants a security interest to B transfers loss that would otherwise have fallen on B upon A's insolvency, to C (junior creditor), a stranger to this contract. The argument that the grant of a security interest is no more than an exercise of A's property rights fares no better, given that the extent of these rights is a matter of what the law allows, and whether the law should allow security interests to be created is precisely what is in need of justification.[115] It might be useful to restate this summary, and the privilege view of security interests, using concepts peculiar to insolvency law. An attempt to improve one's position compared to what it would otherwise be under the insolvency distribution scheme, but which does not bring compensating benefits to those relative to whom this improvement would be brought about, is nothing more than a preference, objectionable on exactly the same grounds as any other.[116]

Now as a rule, a *fixed* charge fulfils the requirements of the privilege view. Its holder deserves its priority position in its debtor's insolvency because the charge (a) helps control financial agency costs, or (b) encourages the chargee to provide new funding to troubled companies which, in the absence of the priority of fixed security, would not have been provided, or (c) both. This is mutually beneficial for all the parties interested in the debtor's undertaking since it lowers the risk of its ending up in insolvent liquidation and thus raises the expected value of the claims of all of its creditors, secured and unsecured.[117]

The floating charge brings no such benefits, however. Because it does not entrench the priority of its holder's claim with respect to the collateral, it can make no contribution to the control by its holder of financial agency costs. And because its priority is highly unreliable *ex ante*, ie at the time at which the charge is created, an offer of a floating charge to a potential lender by a troubled company does not noticeably increase the chances of its receiving new funding.[118] The *distinctive* value-enhancing contributions

[114] See, eg, L Bebchuk and J Fried, 'The Uneasy Case for the Priority of Secured Lending' (1996) 105 *Yale LJ* 857, 933.

[115] See generally A Clarke, 'Re-locating security interests within the property framework', in Harris (ed), *Property Problems from Genes to Pension Funds* (London, Kluwer, 1997).

[116] Even if it does not fall within s 239 of the Insolvency Act 1986, perhaps because of the much-criticised shortcomings of its drafting (notably, the focus on the debtor's 'desire' to prefer).

[117] This argument and supporting empirical evidence can be found in detail in Mokal, 'The Search for Someone to Save: A Defensive Case for the Priority of Secured Lending' (2002) 22 *OJLS* 687.

[118] The new legislation's linking of the right unilaterally to appoint an administrator increases the inducement for a creditor to include a floating charge in its debenture when otherwise it might not have. It does not create a significant inducement for the making of a loan on the security of a floating charge when the loan would not otherwise have been made, since—in the

of the floating charge have already been discussed, and while these contributions persisted, it was perhaps justifiable for its holder to be given the benefit of its priority in the small proportion of insolvencies where there would be such a benefit.[119] However, it has been argued that the floating charge no longer performs any distinctive socially useful role. All the social benefits previously derived from coupling it with fixed security and receivership can now be captured simply through the combined operation of fixed security and the new administration procedure.

Nevertheless, commercially powerful lenders (most obviously, banks) would continue to include floating charges in their debentures because (a) it is statutorily required in order to obtain the ability unilaterally to appoint an administrator (though nothing special about its nature makes this necessary), and (b) they would have nothing to lose and some advantage to gain from the prospect, however distant, of the usefulness of its priority. However, this priority, no matter how unlikely it is to be efficacious, would be merely exploitative when it *did* turn out to be useful, accomplishing nothing except the improvement of the position of its holder compared to what it would otherwise have been. In those rare cases, it would be draining insolvency value away from ordinary unsecured creditors, without (any longer) bringing them compensating benefits. Simply put, the retention of the floating charge now creates a mechanism for preferences to be obtained in some insolvencies which are every bit as objectionable, as a matter of principle, as any caught and avoided by section 239 of the Insolvency Act.[120]

VII. SOME CONCLUDING THOUGHTS

This paper has argued that the usual way of conflating floating with fixed charges as small variations on a single theme—as priority-based devices differing only in degree—fundamentally misunderstands its true nature. The floating charge has played a distinctive role as a residual management displacement device which could only be effective if coupled with an appropriate set of fixed security that enabled its holder to gather information about the competence of the debtor's managers and to control their incentives to misbehave. The floating charge allowed the debtor free use of

absence of the benefit of some priority-based security—the right to appoint an administrator brings little *private* advantage to the appointor.

[119] Put differently, it is being claimed that all the relevant parties bargaining in the choice position of the Authentic Consent Model would arguably have found the relatively small expected costs of allowing floating charge holders priority in a small proportion of insolvencies to be outweighed by the expected mutual benefits of the floating charge described above; see Mokal, 'The Authentic Consent Model' [2001] *LS* 400.

[120] Note a qualification: some preferences might be objectionable because they destroy synergies by removing an important asset from the debtor's estate for the benefit of the preferred creditor. This point is not implicated in the text here. What matters is that because of

its circulating assets while its management was doing well, and when the management failed, it would crystallise to divest them of control over these assets. At this time, it would also contribute to controlling the motivation costs of other creditors by discouraging them from rushing to enforce their claims against circulating assets or threatening to do so, and by ensuring the unity of the debtor's estate under the receiver's control even after the onset of winding-up. This priority-independent view of the floating charge enabled us to explain the peculiarities of its actual operation, to understand why it was generally taken in combination with fixed charges, and either over the debtor's entire undertaking or over those assets not amenable to the draconian control of the fixed charge, and also to bring out the non-anomalous status of the 'lightweight' floating charge. The analysis also helped throw light on the role of registration, and on the problems occurring at the boundary between the two types of charge. Despite all this, however, it was argued that with the onset of changes introduced to the administration regime by the Enterprise Act, the floating charge has outlived its utility, and that its continued retention is now difficult to justify.

It would be useful at this point to highlight three important issues that have not been examined in this paper. First, and even if the arguments developed here do explain the nature of the floating charge as it exists at the present time, do they apply to its early history as well?[121] Has the floating charge *never* been a priority-based device? While I am content with the claim that the above analysis explains the floating charge as it operates at the present, there is at least some reason to think that it would never have been relied upon primarily for its priority, both because it would never have encumbered the collateral, and because it was always subject to preferential and (when there was a distinction between the two) Crown claims[122] which would be difficult to quantify *ex ante*. Second, it is important to note that the points made above have significant implications for the validity of 'automatic crystallisation clauses', which purport to crystallise a floating charge upon the occurrence of a specified event even when neither of the parties is aware of this. Once again, this is an effort to acquire priority for the holder of the floating charge without requiring it to exercise control over the collateral. For the reasons explained above,[123] the efficacy of this must be in serious doubt, normatively *and* by implication of the Privy

the priority of the floating charge, the preferred creditor would receive value that ought to have gone to other creditors.

[121] That there might be such a distinction was suggested by Jacob Ziegel at the Commercial Law and Commercial Practice seminar at the LSE in November 2002 in response to my argument that the floating charge might not be a priority-based device, and has been reiterated since by Michael Bridge and Ian Fletcher.

[122] An excellent historical overview can be found in A Keay and P Walton, 'The Preferential Debts Regime in Liquidation Law: In the Public Interest?' [1999] *CfiLR* 84, 86–91.

[123] See Sections IV and VI above.

Council's decision in *Agnew*. And finally, space does not allow consideration of whether and how the analysis here fits in with the Law Commission's recent proposals about the reform of security interests.[124] These questions, and others, require further thought.

I want to conclude by quickly suggesting two ways in which we might deal with the problem that, as I claimed above, has now arisen from the retention by the Enterprise Act of the floating charge. First and more obviously, the floating charge might be abolished outright, perhaps as part of the reforms which, it is hoped, would be brought about as a result of the Law Commission's ongoing investigation into reform of security interests. Alternatively, the Enterprise Act has laid down that a certain proportion of the value of property subject to a floating charge is to be distributed to unsecured creditors.[125] The appropriate proportion has yet to be specified. Now if the floating charge were a priority-based device, then this restriction on its priority might well have been objectionable, since, for example, it would dilute the ability of a troubled firm to attract new funding, thus harming all those with claims against it.[126] However, in view of the arguments made above, it is submitted that there is a strong case for setting at the highest level politically feasible the proportion of property which should go to unsecured creditors, right up to one hundred percent. The floating charge is not relied upon for priority, except opportunistically (ie *ex post*), so this would result in some benefit to unsecured creditors and would cause no harm to creditors who take security.

[124] See *Registration of Security Interests: Company Charges and Property other than Land* (Consultation Paper 164, London, TSO, June 2002).

[125] See s 176A of the Insolvency Act 1986, as inserted by s 252 of the Enterprise Act.

[126] See eg, Mokal, 'Search for Someone to Save', esp at 721–7.

Part 5

Controlling Modern Management

18

Contractual Modification of the Duties of a Trustee

MICHAEL BRYAN

I. INTRODUCTION*

LAWYERS ARE FAMILIAR with commercial applications of the trust.[1] Secured lending, managed investments and property schemes can all be wholly or partially structured through trust machinery. Motivations for adopting this structure in preference to available alternatives are various. The 'creditor proofing' of transactions is an obvious stimulus. Where *Quistclose* trusts[2] or reservation of title clauses incorporating trusts[3] are employed the 'proofing' is directed not just at the immediate borrower or buyer, but at any floating charge imposed by a bank over that party's assets.[4] The trust's fiduciary regime is also an attraction, especially to legislatures who have recognised that it provides a rigorous but flexible 'off-the-peg' ethical regulatory framework, capable of filling gaps in even the most detailed legislative scheme.[5] There may also be tax advantages associated with the adoption of a trust structure, although a discernible if uneven shift towards 'entity taxation', eliminating the differences attributable to the selection of a specific legal structure, may in the long term remove the significance of fiscal benefits.[6]

* See now Law Commission, Consultation paper 171, Trustee Exemption Clauses, especially part 4, for an analysis of the issues considered in this section.

1 John H Langbein, 'The Contractarian Basis of the Law of Trusts' (1995) 105 *Yale Law Journal* 625; Graham Moffat, *Trust Law Text and Materials,* 3rd edn (Butterworths, 1999), chs 13–16; Michael Bryan, 'Reflections on Some Commercial Applications of the Trust' in Ian Ramsay (ed), *Key Developments in Corporate Law and Trusts Law: Essays in Honour of Professor Harold Ford* (2002) ch 9.

2 *Barclays Bank Ltd v Quistclose Investments Ltd* [1970] AC 567.

3 *Associated Alloys Pty Ltd v ACN 001 452 1026 Pty Ltd* (2000) 202 CLR 588.

4 *Clough Mill Ltd v Martin* [1984] 3 All ER 982.

5 *Pensions Act* 1995 (UK); *Superannuation Industry (Supervision) Act* 1993 (Cth) (Aus).

6 *A Tax System Redesigned: Review of Business Taxation* (Ralph Committee Report, 1999) 472–481, 546–551.

The 'flexible interplay of law and equity'[7] is a characteristic feature of commercial trust applications, although in areas such as pensions and other managed investments a layer of statutory regulation has been superimposed on the private law concepts. Lord Browne-Wilkinson drew attention to the specifically contractual context of many commercial trusts in *Target Holdings Ltd v Redferns* when suggesting that not all 'traditional trust' principles will be appropriate to the enforcement of a trust which has been constituted as machinery for effectuating contractual objectives.[8] Recent cases have explored the application of contractual techniques and policies to aspects of express trusts. They include analysis of the circumstances in which equity should 'follow the law' in awarding compensation for breach of an express trust,[9] the implication of terms into trust arrangements,[10] and the impact of public policy where entry into a contract or transfer of property infringe a legislative prohibition.[11]

It is tempting to announce the arrival of the 'contractualisation' of the commercial trust as a legal phenomenon, but to do so would obviously be premature. Trusts created for commercial objectives are too diverse, in form as well as function, to permit any sensible generalisation as to the relative significance of contract and trust principles, to be formulated. Some beneficial interests under such trusts resemble a 'half life', for example money held on loan pending finalisation of the loan, and the arrangement of collateral security, when the trust will be superseded by contractual and proprietary protection for the lender. In others, such as collective investment schemes, the trust constitutes an enduring feature of the scheme and trust principles, such as the rule in *Saunders v Vautier*,[12] confer important protection on the beneficiary.[13] The scale and intensity of trustees' duties will also vary. The bare trustee, whose principal duty is simply 'to act in accordance with instructions', has clearly taken on responsibilities of a different order than those imposed on the trustee of a collective investment trust. The latter will be subject to all the equitable and Trustee Act duties of a trustee, save to the extent that their application has been modified or excluded by regulatory legislation. The incidence of application of contractual principles and methodology to commercial trusts varies not only between different types of trust, but also between trusts of the same generic description.

This paper examines one example of the application of contractual principles to trusts: that of construing exemption clauses (or exculpation clauses,

[7] *Barclays Bank Ltd v Quistclose Investments Ltd* [1970] AC 567, 582, per Lord Wilberforce.
[8] [1996] 1 AC 421, 435.
[9] *Canson Enterprises Ltd v Boughton & Co* (1991) 85 DLR (4th) 129.
[10] *Associated Alloys Pty Ltd v ACN 001 452 1026 Pty Ltd* (2000) 202 CLR 588.
[11] Compare *Tinsley v Milligan* [1994] 1 AC 340 with *Nelson v Nelson* (1995) 184 CLR 538.
[12] (1841) Cr & Ph 240; 41 ER 482.
[13] Michael Bryan, 'Reflections on Some Commercial Applications of the Trust' in Ian Ramsay (ed), *Key Developments in Corporate Law and Trusts Law: Essays in Honour of Professor Harold Ford* (2002) ch 9.

as they are commonly known) in trust instruments. Clauses excluding the liability of trustees for breach of trust, either generally or with savings for dishonesty or gross negligence, have a long history in books of precedents.[14] Until recently, however, authority on the construction and enforcement of these clauses was sparse.[15] The Court of Appeal decision in *Armitage v Nurse*[16] has now filled the gap in authority by laying down the principles of construction applicable to exculpation clauses. It has been applied, refined and distinguished in later English and Commonwealth decisions which have in turn shed light on the operation of these principles. Millett LJ adopted an essentially contractual approach to the construction of exculpation clauses, while recognising that an 'irreducible core' of fundamental equitable obligation is not amenable to exclusion.[17] This paper assesses, with reference to cases decided since *Armitage v Nurse*, the appropriateness of applying contractual techniques developed in the context of contractual exclusion, to trustee exculpation. It will be suggested that defining the limits of trustee exclusion is not wholly analogous to construing the proper scope of a contractual exemption clause, and that the 'irreducible core' of trust obligation may be larger than envisaged by Millett LJ in *Armitage v Nurse*.

II. THE FOUNDATION CASE: *ARMITAGE v NURSE*

The plaintiff was a beneficiary under a family trust which included land farmed by a family company. The plaintiff's mother and grandmother were sole directors and shareholders of the company. The plaintiff alleged a number of breaches of trust, including the trustees' failure to supervise the company's management of the land, and failure to obtain proper payment of interest of a loan of trust money to the plaintiff's mother. The principal issue was whether an exemption clause absolved the trustees from liability for any or all of the breaches. The widely drawn clause provided that: 'No trustee shall be liable for any loss or damage which may happen to [the plaintiff's] fund or any part thereof at any time or from any cause whatsoever unless such loss or damage shall be caused by his actual fraud.' Delivering the judgment of the Court of Appeal Millett LJ held that the

[14] Hallett's *Conveyancing Precedents* (1965); Key and Elphinstone's *Precedents in Conveyancing* 15th edn, (1953).
[15] HAJ Ford and WA Lee, *Principles of Law of Trusts* 2nd edn, (1990) noted (para [1806]) the 'dearth of authority' on this subject. Contrast the wealth of authority discussed in the 3rd ed, (September 1998) para [18060]. A feature of United Kingdom law, until the Court of Appeal decision in *Armitage v Nurse* [1998] Ch 241, was the scarcity of English authority compared with the plenitude of Scottish cases on the subject. See *Seton v Dawson* (1841) 4D 310; *Knox v Mackinnon* (1888) 13 App Cas 753; *Rae v Meek* (1889) 14 App Cas 558; *Wyman v Patterson* [1900] AC 271; *Clarke v Clarke's Trustees* (1925) SC 693.
[16] [1998] Ch 241.
[17] *Ibid*, 253–254.

clause afforded the trustees complete immunity from liability for the alleged breaches. The reference to 'actual fraud' connoted dishonesty on the part of the trustees; it did not encompass 'equitable fraud', in the sense of a breach of fiduciary obligation or the commission of some other wrongdoing.[18] In addition to dishonest breaches of trust, the exculpation clause was not effective to exclude liability for two other types of breach:

1. Cases where the beneficiary was not suing in respect of any 'loss or damage' to the trust, within the terms of the clause.[19] This was said to include the purchase of trust property by a trustee. A claim for the restitution of trust property, or of its value, is not a claim to compensation for loss or damage to the trust. Millett LJ reached this conclusion by reference to the wording of the exculpation clause in question. It is also arguable that a beneficiary's right to compel restitution and reconstitution of the trust fund is a core trust obligation not capable of exclusion.
2. A breach of duty going to the 'irreducible core of obligations owed by the trustees to the beneficiaries and enforceable by them' and 'which is fundamental to the concept of a trust'.[20] The 'irreducible core' was defined restrictively, so as to give the broadest possible scope to the principles of contractual construction:

   "If the beneficiaries have no rights enforceable against the trustees there are no trusts. But I do not accept the further submissions that these core obligations include the duties of skill and care, prudence and diligence. The duty of the trustees to perform the trusts honestly and in good faith for the benefit of the beneficiaries is the minimum necessary to give substance to the trusts, but in my opinion it is sufficient."[21]

It followed that the clause was effective to exclude liability for breach causing loss or damage to the trust fund, not being a breach of a core non-excludable obligation.

Millett LJ's construction of the exculpation clause applied the well known principles of contractual exclusion. While an exculpation clause will be construed strictly against the interest of a trustee seeking to rely upon its terms, it will also be presumed that the clause was intended to cover at least some breaches, and not to be mere surplusage. The analogy between

[18] *Ibid*, 250–251.
[19] *Ibid*, 253.
[20] *Ibid*, 253–254.
[21] *Ibid*.

excluding liability for common law negligence and exempting liability in equity for want of business prudence was explicitly noted:

> It is, of course, far too late to suggest that the exclusion in a contract of liability for ordinary negligence or want of care is contrary to public policy. What is true of contract must be equally true of a settlement.[22]

Even the exceptions to the application of contractual principles of construction are more apparent than real. The refusal to apply an exculpation clause to breaches not causing loss or damage mirrors the insistence at common law upon the principle that an exemption clause can only apply within the 'four corners' of the contract, and is therefore ineffective to exclude liability caused by breaches which cannot be construed as coming within the terms of the clause.[23] Moreover, the irreducible core obligation of trusteeship which cannot be excluded can be regarded as the equitable counterpart of the contractual principle that no clause can authorise performance of an obligation wholly different from that envisaged by the contract. The venerable textbook example is that of a seller of beans who delivers peas to the purchaser.[24] No exclusion clause, however widely drawn, can protect the seller from liability for failing to perform the contract at all. All legal obligations which facilitate the transfer of wealth have a non-excludable essence in order to preserve the principle that a legal, and not simply a moral, obligation has been imposed.

III. SOME POLICY ARGUMENTS

Should the principles applicable to contractual exclusion also govern trustee exculpation? An obvious rejoinder is to ask: what other principles could possibly be applied? Although trusts law and the law of succession enjoy their own principles of construction they do not include principles which deal with trustee or executor immunity from liability. Contract law possesses the most fully realised techniques for construing exception clauses, as well as a rich corpus of case law illustrating the application of these techniques. It would be perverse for a trusts lawyer to ignore the established contractual approaches for excluding and limiting liability. Moreover, as has already been noted, some trusts are created as part of the enabling machinery of an essentially contractual arrangement. In these cases the principles governing the enforceability of an exculpation clause to protect a contracting party acting as trustee ought logically to be consistent with the principles determining the enforceability of a contractual exemption clause.

[22] *Ibid*, 254.
[23] Treitel, *The Law of Contract* 10th edn, (London: Sweet & Maxwell, 1999), 207–211.
[24] *Chanter v Hopkins* (1838) 4 M&W 399, 404; *George Mitchell (Chesterhall) Ltd v Finney Lock Seeds Ltd* [1983] 2 AC 803, HL.

These are compelling reasons for drawing on the resources of contract law in interpreting a trust instrument. But they can be counterbalanced by other reasons for not extending the contractual methodology to trusts. Some of these considerations are of a general nature; others are specific to discrete trust applications:

1. Where property is settled by a settlor on trust the beneficiaries are third parties to the 'deal' negotiated between settlor and trustee.[25] While legislation in common law jurisdictions has removed many of the obstacles to enforcement by third parties to a contract, the application of an exemption clause to which the beneficiaries have not agreed raises different considerations from the application of such a clause to which a contracting party has objectively agreed, however artificial that objective agreement may be in practice.

   An obvious distinction can be drawn in this connexion between volunteer beneficiaries and beneficiaries who have provided consideration as part of a commercial trust arrangement. It is plausible to argue that an exculpation clause should more readily be applied to defeat a claim brought by a beneficiary who has not paid for her interest than one who has.[26] But drawing this distinction oversimplifies the question of enforcement. Where the beneficiary has subscribed to a managed investment scheme, such as a unit trust, the beneficiary has, for the purposes of objective will theory, 'chosen' to buy an equitable interest under a trust with an instrument which excludes or limits the trustee's liability.[27] The exercise of this choice is an argument for applying the exemption clause to defeat any claim for breach of trust brought by the beneficiary.
2. Exemption and exclusion clauses immunise a party in default from *personal* liability for breach of trust. Unlike a breach of contract, a breach of trust may entitle the beneficiary to proprietary relief as an alternative to personal recovery. Equitable proprietary orders restore to beneficiaries interests which, 'ex hypothesi', already belong to them.[28] Just as courts cannot excuse a trustee who has

[25] This analysis obviously does not apply to cases of 'self declaration' by a settlor constituting herself as trustee. See above n 1, John H Langbein, 'The Contractarian Basis of the Law of Trusts' (1995) 105 *Yale Law Journal* 625, 672–675.
[26] Compare *Armitage v Nurse* [1998] Ch 241 with *Reader v Fried* [2001] VSC 495.
[27] The artificiality of applying will theory to the purchase of interests under a managed investment is one reason why legislation has imposed 'non-excludable' obligations on trustees, especially pension and superannuation trusts. See *Pensions Act* 1995 (UK ss 32–36); *Superannuation Industry (Supervision) Act* 1993 (Cth) s 52.
[28] This would not, of course, be the case if a legal system were to grant remedial and redistributive proprietary interests not based on a claimant's pre-existing entitlement. See the debate generated by the publication of D Wright, *The Remedial Constructive Trust* (Sydney, Butterworths, 1998).

committed an honest breach of trust from the duty to restore to the trust specific property which he has wrongly acquired and retained,[29] a trustee cannot justify misappropriation of another's property by reference to an exemption clause, not being a provision effective to transfer the interest to him.[30] This could be rationalised in terms of the trustee's obligation not to misappropriate trust property being, in Millett LJ's phrase, a 'core irreducible obligation' of trusteeship. And so it is. But just as it strains legal analysis to explain property law solely in terms of the duty not to commit theft, so it would be a distortion to rationalise the beneficiary's right to compel the restitution of specific trust assets in terms of a breach of a trustee's duty not to misappropriate. Unless the trustee has some justification for beneficially retaining the trust property, for example as a trustee-beneficiary, an exculpation clause cannot bar a beneficiary's claim to proprietary recovery.

3. Whereas remedies for breach of trust focus on the disgorgement of unauthorised gains at least as often as upon compensation for loss, the primary objective of monetary remedies for breach of contract is compensatory, being to place the plaintiff in the position in which she would have been if the breach had not occurred. Recent authority has converted this difference from being one of principle to being one of degree. An account of profits can now be awarded for a breach of contract.[31] But the award of the remedy in this context remains, to borrow a famous fifteenth century dictum, a 'new found halliday',[32] and the precise circumstances in which accounting will be ordered are contentious. However the matter is ultimately resolved, it can confidently be stated that accounting for gains will remain the exception to the rule that expectation or reliance based damages are the standard monetary awards for breach of contract.

   Compelling a trustee or other fiduciary to disgorge gains made through a breach of obligation is more characteristic of equity. Most exculpation clauses absolve the trustees from liability for 'loss or damage' caused by a breach of duty and therefore confer no protection against unauthorised profit making. How a clause, which in terms prevents beneficiaries, or a successor trustee acting on behalf of the beneficiaries, from reclaiming the profit will be construed can only be a matter for speculation. Cases decided after *Armitage v Nurse* have been reluctant to give effect to

[29] *Trustee Act* 1925 s 61. John Lowry and Rod Edmunds, 'Excuses' in Peter Birks and Arianna Pretto (eds), *Breach of Trust* (Oxford, Hart, 2002) 269.
[30] Compare James Penner, 'Exemptions' in *Breach of Trust*, above n 29, 265–267, analysing a proprietary exemption clause as a contingent interest in the property itself.
[31] *Attorney-General v Blake* [2001] 81 AC 268.
[32] Y B 33 Hen V1, f 26, pl 12, Littleton J.

exculpation clauses, as a matter of construction, where the loss caused by the breach to the trust estate is matched by a corresponding gain to the trustees. It is conceivable that an exclusion clause permitting a trustee to retain gains might be held to be unenforceable on the ground that the profits are no more than the natural augmentation of the trust property, and that trustees are not entitled to immunity from the duty to restore misappropriated trust property.[33] If, as is sometimes argued, the availability of the accounting remedy fulfils a prophylactic purpose in deterring fiduciaries from exploiting the profit-making opportunities inherent in their position, that deterrent function would be furthered by a prohibition on reliance on exculpation clauses where the disgorgement of gains is sought.

But deterrence, applied as an abstract value to justify disgorgement of improper fiduciary gains, is a crude basis for preventing reliance on an exclusion clause. It tends to oversimplify any inquiry into the nature of gains lawfully made by the fiduciary. Some gains may represent remuneration or commission properly claimed by the trustee, particularly where the trustee is payable by commission under a managed investment scheme. In other cases the gain may represent opportunities that the beneficiary could not have taken for herself, entitling the fiduciary in breach to an allowance or remuneration.[34] Even if deterrence is recognised as a policy informing the imposition of equitable relief on defaulting fiduciaries, it is unclear whether it can justify an absolute prohibition on reliance on exculpation clauses, especially where the breach of trust is 'beneficial' in that it enables beneficiaries to enjoy a profit which could not have been made by strict adherence to the letter of the trust instrument.

4. Trustees are entitled to an indemnity for costs and expenses incurred in the performance of their work, whereas contracting parties must bear the costs of their performance unless the contract provides otherwise. The trustees' right of indemnity from the trust fund is inherent in the trust relationship, and probably cannot be excluded by contrary provision in the trust instrument.[35]

[33] *Foskett v McKeown* [2001] 1 AC 102 supports the thesis that profits derived from the misappropriation of trust property constitute the natural increment of that property. But the decision has been criticised for this property analysis. Peter Birks, 'Property, Unjust Enrichment and Tracing' (2001) 54 *Current Legal Problems* 231; Andrew Burrows, 'Proprietary Restitution: Unmasking Unjust Enrichment' 117 (2001) *Law Quarterly Review* 412, 419–423.

[34] *Boardman v Phipps* [1967] 2 AC 46; *Warman International Ltd v Dwyer* (1995) 182 CLR 544.

[35] *Vacuum Oil Co Pty Ltd v Wiltshire* (1945) 72 CLR 319 though exclusion was stated, obiter, to be possible in *Re RWG Management Ltd* (1984) 9 ACLR 739. Compare the right of indemnity from the beneficiaries, which is capable of exclusion and is almost invariably excluded in managed investment trusts: *Hardoon v Bellilios* [1901] AC 119; *McLean v Burns Philp Trustee Co Pty Ltd* (1985) 2 NSWLR 623, 640–1.

> Trustees are entitled to an indemnity even if their actions on behalf of the trust constitutes a tort.[36] They can claim reimbursement for acts performed which constitute a breach of trust, while acting for the benefit of the trust, although they will not usually be entitled to the benefit of their indemnity unless they have first made good to the trust estate the loss occasioned by the breach.[37]
>
> The trustees' right to an indemnity fulfils a different purpose from their reliance upon an exculpation clause. But the existence of the right illustrates an important qualification on treating exculpation clauses in trust instruments as if they were contractual exclusions. If a contractual exemption clause is held not to apply to a breach, then, subject to the availability of other defences and the plaintiff's duty to mitigate, the plaintiff will be entitled to be fully compensated for the consequences of the breach. But if an exculpation clause is held not to cover loss caused by a breach of trust an award of compensation may be offset by a trustee's claim to indemnity for undertaking activities beneficial to the trust. Indemnity is an important feature of trading trusts, and may also be exercised by trustees of managed investment trusts. The existence of the right is a reminder that not all the incidents of trusteeship are explicable in terms of contractual bargain, or are excludable on this basis.[38]

Not all these grounds for distinguishing the contractual methodology for construing exemption clauses have been argued in the post-*Armitage v Nurse* cases on the interpretation of exculpation clauses. Recent authority on the trustee's right to an indemnity is, for example, slight. But other distinctive features of breaches of trust, such as the fact that these breaches result in a higher incidence of unauthorised profit making than breaches of contract, have certainly influenced the construction of these clauses. Cases decided since *Armitage v Nurse* have taken that decision as their starting point, treating it as authority for the proposition that contract rules of construction should be applied to exclusion clauses in trust instruments. But they have rarely followed *Armitage v Nurse* in holding that the clause does, in terms, apply to the breach.[39] The methods applied to distinguish the decision will be considered in the next section.

[36] *Re Raybould* [1900] 1 Ch 199.
[37] *McEwan v Crombie* (1883) 25 Ch D 175; *Jesse v Lloyd* (1883) 48 LT 656.
[38] See also Robert Chambers, 'Liability' in *Breach of Trust*, above n 29, 8–10, for other examples of the trustees' right to enter losses as disbursements on the trust account, such as robbery of trust assets. Law Commission Consultation Paper 171, n* above, para 4.88 recommending that indemnity should not be allowed where reliance on an exemption clause is not permitted.
[39] In addition to the cases discussed in the next section, *Armitage v Nurse* has been applied in *Allan v Rea Brothers Trustees Ltd* [2002] EWCA Civ 85 where an exculpation clause was applied to defeat a claim brought by a beneficiary who had participated in the breach of trust, and in *Bogg v Raper* [1998] EWCA Civ 661 where in the course of making sense of apparently contradictory clauses, the Court of Appeal held that the exculpation clause did not apply to the breaches in question.

IV. LATER DECISIONS

A widely drawn exculpation clause may be held not to extend to a breach of trust for two reasons, which are not mutually exclusive. The first, following *Armitage v Nurse*, is that the clause does not in fact apply to the breach which has occurred. The court is entitled to rely on the contractual principle that the clause should be construed narrowly against the party relying on it. On the other hand, as we have also seen, the clause must also be presumed to have some content and, therefore, to confer protection against some breaches of trust. Where the breach is deliberate there is an understandable temptation to narrow the categories of breach to which the exculpation clause could apply.

The recent decision of the New South Wales Court of Appeal in *Alexander (trading as Minter Ellison) v Perpetual Trustees WA Ltd*[40] illustrates the technique of reading down the scope of exculpation clauses. Money was invested by the plaintiffs, who were approaching retirement, in trusts of which Perpetual was the trustee. The scheme involved investment in a company, ECCC Ltd, who had represented that the scheme was suitable for the plaintiffs. ECCC Ltd in fact invested the plaintiffs' money in the international money markets and in commodities contracts without adequate expertise or resources to undertake this high risk investment strategy. The plaintiffs' money was lost. The plaintiffs successfully sued ECCC Ltd and the trustee. The trustee in turn successfully cross-claimed against the solicitors who had also been involved in the administration of the scheme. One of its key features was that the trustee deposited the investors' money with the solicitors. The solicitors released the money to ECCC upon production of a bearer certificate which could be drawn by an investor at the Dresdner Bank. In executing the scheme the solicitors placed themselves in a position of conflict of interest since they acted as solicitor for the trustee as well as agent for ECCC Ltd in collecting money on the company's behalf. The solicitors acted in breach of the obligations they owed to the trustee by accepting documentation from ECCC Ltd which did not constitute negotiable bearer documents, and which could not be drawn upon the bank by individual investors. The solicitors relied upon an exclusion clause in the subscription agreement which read:

> The parties agree and acknowledge that the Company's solicitors have no liability for disbursement of the funds held pursuant to [the investment agreement] and the parties indemnify and hold the Company's Solicitors harmless for any claim, demand, action or cost arising out of or in connection with the disbursement of any funds under this Agreement.[41]

[40] [2001] NSWCA 240, upholding *Wilkinson v Feldworth Financial Services Pty Ltd* [1998] NSWSC 775; (1999) 17 ACLR 220.

[41] Clause 3.4 of the Subscription Agreement. The clause was reinforced by a sentence in the letter signed by all subscribers: 'Further, I absolutely and unconditionally release and discharge you for any liability or obligation to me under the Subscription Agreement.'

The New South Wales Court of Appeal, upholding the first instance decision of Rolfe J, held that the clause did not apply to the solicitors' failure to obtain bearer securities from ECCC Ltd for the investors. The clause protected the solicitors only if they complied with their responsibilities under the subscription agreement (or, in contractual parlance, if they acted 'within the four corners of the agreement'). For example, it would have protected the solicitors from liability if investors' money had been lost after the correct bearer certificates had been obtained and the money released to ECCC Ltd, or if money had been released to enable ECCC Ltd to fund the issue of the certificates and later lost.

Applying the narrow approach taken by the Court of Appeal, the clause could never apply to a breach of fiduciary duty occurring when the money was held by solicitors, but could operate to defeat a claim of fiduciary duty where the loss had occurred after the money had been released in accordance with the terms of the subscription agreement. The decision constitutes a strict, though not unreasonable, application of the '*contra proferentem*' principle of construction. Although a conflict of interest is not an 'irreducible core obligation' of trusteeship,[42] the Court was understandably reluctant to apply the clause to a breach which arose out of a concealed conflict of interest. Individual investors could not have realised that the solicitors were acting as agent for the company receiving their money as well as on their behalf. They had not seen the documentation appointing the solicitors as the agent of the trustee as well as of ECCC Ltd. The undisclosed conflict of interest is a type of breach more characteristic of trusts law than contract law, where, whatever the problems in construing the exception clause, the parties are usually recognised as possessing commercially discrete interests. Will theory holds that parties have objectively consented to the agreements they sign, and provides only very limited relief for material non-disclosure. But *Alexander v Perpetual Trustee* demonstrates that these contractual presuppositions have little application where a beneficiary's losses have been caused by a concealed and undiscoverable conflict of interest.

The second technique employed to prevent reliance on an exculpation clause is to find that a trustee has acted dishonestly. Millett LJ in *Armitage v Nurse* held that the reference to 'actual fraud' in an exemption clause 'means what it says'.[43] But the construction of 'dishonesty' in trust instruments has not proved to be quite as straightforward. Analyses of the term

[42] *Kelly v Cooper* [1993] 3 AC 205. A trustee may of course also be a beneficiary though the exercise of a discretion may require the trustee to surrender the exercise of discretion to the court: *Re Drexel Bumham Lambert UK Pension* [1995] 1 WLR 32, Ch cf *Pensions Act* 1995 s 39.

[43] *Armitage v Nurse* [1998] Ch 241, 250.

have applied the objective meaning of this concept, elaborated by Lord Nicholls in *Royal Brunei Airways v Tan*:

> This is an objective standard. At first sight this may seem surprising. Honesty has a connotation of subjectivity, as distinct from the objectivity of negligence. Honesty, indeed, does have a strong subjective element in that it is a description of a type of conduct assessed in the light of what a person actually knew at the time, as distinct from what a reasonable person would have known or appreciated. Further, honesty and its counterpart dishonesty are mostly concerned with advertent conduct, not inadvertent conduct. Carelessness is not dishonesty. Thus for the most part dishonesty is to be equated with conscious impropriety. However, these subjective characteristics of honesty do not mean that individuals are free to set their own standards of honesty in particular circumstances. The standard of what constitutes honest conduct is not subjective. Honesty is not an optional scale, with higher or lower values according to the moral standards of each individual. If a person knowingly appropriates another's property, he will not escape a finding of dishonesty simply because he sees nothing wrong in such behaviour.[44]

Lord Nicholls was considering the meaning of dishonesty for the purposes of imposing liability for dishonest assistance in a breach of fiduciary duty. But this passage has also influenced the interpretation of dishonesty in the context of a claim by a beneficiary that a trustee is precluded from relying on the terms of an exemption clause by reason of his dishonesty. The clauses in the exemption clause cases have invariably included savings for fraud or dishonesty in what are otherwise comprehensive exclusions, but even if a clause were expressly to exclude liability for a dishonest breach of trust, a court would never permit it to apply to such a breach since the duty of good faith is part of the irreducible core of trusteeship.

In exemption clause cases objective dishonesty can be identified by reference either to the status of the trustee or to the nature of the breach. Taking status-based dishonesty first, solicitor trustees have been held to be dishonest in circumstances in which non-professional trustees do not have dishonesty imputed to them. *Alexander v Perpetual Trustees WA Ltd* is an example of judicial willingness to characterise an advertent breach of trust committed by a trustee as dishonest. An even clearer example is the Court of Appeal decision in *Walker v Stones*.[45] The trustees of a family discretionary trust were solicitors. The principal trust assets were the two issued shares in a parent company which owned a controlling interest in another company. The trustees arranged for that company to subscribe to bonds in a third company, owned by the settlor of the trust who was also the father of the beneficiaries. The bonds were worthless, with the result that the entire net value of the trust assets were dissipated by the investment. In answer to a claim of breach of

[44] [1995] 2 AC 378, 389.
[45] [2001] 2 WLR 623, at the time of writing on leave to appeal to the House of Lords.

trust brought by the beneficiaries the trustees relied upon an exemption clause which in its essentials provided that:

> In the professed exercise of the trusts and powers hereof no trustee ... shall be liable ... by reason of any ... matter or thing other than wilful fraud or dishonesty on the part of the trustee.

The Court of Appeal held that the pleadings established a sufficient basis for an allegation of dishonesty, based on inferences drawn from the conduct of the trustees that they had consistently exercised their powers to further or protect the financial interests of the settlor-father rather than the interests of the children-beneficiaries. In construing the proviso for 'wilful fraud or dishonesty' in the exemption clause, Sir Christopher Slade applied the test of 'objective dishonesty' formulated by Lord Nicholls in the *Tan* case. Dishonesty, for this purpose, included a belief that was so unreasonable that no reasonable solicitor-trustee could hold it.[46]

An attraction of the 'dual' standard of honesty, whereby the beliefs of solicitors and other professional trustees may be held to be more unreasonable than the beliefs of lay trustees, is that it is no more than the natural corollary of the 'dual' standard of business prudence, whereby paid professional trustees are held to a higher standard of business prudence than the standard applied to unpaid lay trustees.[47] It is well known that the impact of the higher standard of care imposed upon professional trustees and trust corporations can easily be nullified by the exercise of commercial bargaining power by such trustees to impose standard form trust obligations, including exculpation clauses of superabundant width, upon settlors and (where beneficial interests are purchased) beneficiaries. These are the trust lawyer's equivalent to the contract lawyer's adhesion contract. The turbulances in bargaining power attributable to the relatively small oligopolistic market of professional trustees will be reduced if dishonesty carries the wider signification proposed by Sir Christopher Slade.[47a]

But the application of an objective standard of dishonesty also has its drawbacks. The imposition of 'special' standards of honesty on solicitor-trustees carries the risk, against which Lord Nicholls warned in *Tan*, that the link between dishonesty and its foundation in conscious impropriety be broken. In formulating the standard of honesty for solicitor-trustees Sir Christopher Slade relied upon the Court of Appeal decision in *Twinsectra Ltd v Yardley*[48] where a special standard of honesty applicable to the liability of solicitors as accessories to a breach of fiduciary duty had also been recognised. The Court of Appeal has recently been overruled.[49] The majority

[46] *Walker v Stones* [2001] 2 WLR 623, 624.
[47] *Bartlett v Barclays Bank Trust Co Ltd* [1980] Ch 515, 531–534, Brightman J.
[47a] Law Commission Consultation Paper No 171, n*, para 3.79
[48] [1999] Lloyd's Rep Bank 438, 464, Potter LJ.
[49] *Twinsectra Ltd v Yardley* [2002] 2 All ER 377, 393, Lord Millett dissenting.

House of Lords judgments in *Twinsectra* emphasise that the test of dishonesty laid down by Lord Nicholls combines elements of subjectivity and objectivity. A status based test of dishonesty is open to criticism on the ground that it emphasises the objective element of dishonesty at the expense of its subjective primary meaning deliberate wrongdoing. Lord Hutton expressly drew attention to the risks posed to solicitors by the adoption of a status-specific standard:

> A finding that a defendant has been dishonest is a grave finding, and it is particularly grave against a professional man, such as a solicitor. Notwithstanding that the issue arises in equity law and not in a criminal context, I think that it would be less than just for the law to permit a finding that a defendant had been 'dishonest' in assisting in a breach of trust where he knew of the facts which created the trust and its breach but had not been aware that what he was doing would be regarded by honest men as being dishonest.[50]

Although this dictum is directed to the imposition of personal accountability on solicitors who assist in the commission of a breach of fiduciary duty it also serves as a warning against 'over-objectification' of the standard of dishonesty in the context of trustee exclusion clauses.

The *Tan* standard has also been applied so as to prevent reliance on an exculpation clause, as being dishonest, where a trustee has made a profit from the breach of duty. Here it is the nature of the breach, rather than the identity of the trustee, which attracts the application of the mixed subjective-objective standard favoured by Lord Nicholls. We have already noticed that unauthorised profit-making is more characteristic of breaches of trust than breaches of contract (which is not, of course, to assert that a breach of trust will usually result in the making of an improper profit, or that breaches of contract will invariably cause loss to the innocent party compensatable in damages). Standard form exclusion clauses rarely confer immunity from accounting for profits. This is simple prudence on the part of drafters of these clauses, since any attempt to legitimate unauthorised profit making in this way would set off an alarm bell, at any rate in a commercial trust, sounding to a prospective purchaser of a beneficial interest the fact that the trustee saw no objection to exploiting her fiduciary position for personal advantage. But even if a literal construction of a clause would allow a trustee to make gains not otherwise authorised by the trust instrument, courts may reject this construction on the ground that unauthorised profit making is in its nature dishonest.

The recent Victorian decision of *Reader v Fried*[51] illustrates this example of the 'objectification' of dishonesty. The defendants, husband and wife, were directors of a trustee company which administered a superannuation fund

[50] *Ibid*, 388.
[51] [2001] VSC 495.

on behalf of employees and former employees of the defendants' business. The defendants sold a property, which was no longer needed for their business, to the trustee company. The price paid by the trustee company for the property was $600,000, being $400,000 more than the directors had paid for the property seven years previously, and significantly more than the current market valuation. The trust fund had to borrow money at interest to buy an asset which was of declining importance to the defendants' own business. The transaction placed the defendants in a position of clear conflict of interest and enabled them to make a profit at the expense of the trust. The directors of the trustee company invoked an exculpation clause which read:

> No person being a Trustee or being a member of the Board of Directors of the Trustee or being a director or officer of the Trustee or any other person authorised by the Trustees or the said Board of Directors to act in respect of the Fund shall be under any personal liability in respect of any loss or breach of trust relating to the Fund unless the same shall have been due to his own dishonesty.

Pagone J held that the directors were not entitled to the protection of the clause. Quoting the passage from Lord Nicholls's judgment in *Tan*, extracted above, he concluded that they were dishonest in arranging a sale of property on terms which ensured a profit for themselves and a loss for the trust fund's beneficiaries. It was no answer to the charge of dishonesty that the sale had been negotiated and completed by professional advisers since '[f]iduciaries may rely upon others in the discharge of their duties, but they cannot, without more, shed the duties they have assumed.'[52]

Is *Reader v Fried* open to criticism on the ground of 'over-objectification' to which the House of Lords took exception in *Twinsectra v Yardley*? There are a small number of cases in which a trustee has made a profit, believing on reasonable grounds that she was acting in the best interests of the beneficiaries. *Boardman v Phipps*[53] is an example, though whether the proper response to such a case is to apply an exculpation clause to the breach (had an appropriately worded provision been included in the trust instrument in that case) or to compel disgorgement while permitting the trustee a 'liberal allowance' (the relief actually ordered in *Boardman v Phipps* itself) is open to debate. But such cases aside, the characterisation of improper profit-making as 'dishonest' seems apt. Certainly the characterisation will not result in the grave professional consequences identified by Lord Hutton in *Twinsectra* as being a defect of the status-based definition of dishonesty. Nor will a finding of dishonesty be divorced from the subjective meaning of dishonesty, on which stress was laid by the majority in

[52] *Ibid*, [17].
[53] [1967] 2 AC 46.

Twinsectra, if the trustees, as in *Reader v Fried*, rely on professional advisers and agents to complete the profit-making transaction. Even if the trustees did not will the precise means of making the profit they had willed (and taken advantage of) the end.

V. CONCLUSION

In the words of the venerable legal jest, *Armitage v Nurse* is in danger of becoming a 'very distinguished' case even though it was decided relatively recently. A combination of a strict application of the '*contra proferentem*' principle of construction and an objective interpretation of dishonesty have ensured that exculpation clauses are rarely held to apply to breaches of trust. The literal construction of such a clause, applied in *Armitage v Nurse*, has given way to construction techniques and public policy approaches which have largely preserved the availability of equitable compensatory and profit disgorgement relief.

Consistent with traditional common law (and equitable) analytical methods of reasoning, the case law has been reactive, responding to particular instances of breach of trust with little consideration of the policies that might inform the application of an exculpation clause to a breach. For example, the strongly, indeed viscerally, held view that such clauses should not apply to unauthorised profit making by trustees needs a clearer justification, perhaps in terms of promoting the prophylactic objectives of fiduciary law, than the cases have so far been able to provide. It is not sufficient to state that contractual principles of construction, can automatically be extended to exclusion clauses in a trust deed. Clearly they can—but only up to a point. What is required is an analysis of the contractual ordering, as it applies to trusts law. If *Armitage v Nurse* has shown that, in the context of exculpation clauses, a trust is a contractual 'deal,'[54] later case law has demonstrated that the 'deal' is selective; it applies only for some purposes and to some trustees.

[54] See above n 1, John H Langbein, 'The Contractarian Basis of the Law of Trusts' (1995) 105 *Yale Law Journal* 625.

19

Relieving Directors' Breaches of Duty

ROD EDMUNDS AND JOHN LOWRY*

I. INTRODUCTION

AN ERRANT DIRECTOR may seek total or partial relief from liability under section 727 of the Companies Act 1985 provided the breach of duty is honest and reasonable and the court decides that it ought fairly to be excused.[1] Throughout its century long existence,[2] this judicial relieving provision has failed to excite a great deal of academic discussion or comment.[3] Nor, unsurprisingly, did it figure prominently in the DTI's Company Law Steering Group's deliberations that culminated in sweeping

*This chapter was first published in the *Modern Law Review* ('The Continuing Value of Relief for Directors' Breach of Duty' (2003) 66 MLR 195). The authors wish to record their gratitude to Vanessa Finch and James Penner, both of the LSE, Philip Rawlings, University of Warwick, and to the participants of the LSE Seminar in which this paper was presented for their helpful comments.

[1] S 727(1) provides: 'If in any proceedings for negligence, default, breach of duty or breach of trust against an officer of a company or a person employed by a company as auditor (whether he is or is not an officer of the company) it appears to the court hearing the case that that officer or person is or may be liable in respect of the negligence, default, breach of duty or breach of trust, but that he has acted honestly and reasonably, and that having regard to all the circumstances of the case (including those connected with his appointment) he ought fairly to be excused for the negligence, default, breach of duty or breach of trust, that court may relieve him, either wholly or partly, from his liability on such terms as it thinks fit.'

[2] See the Companies Act 1907, s 32, which was substantially reproduced in the Companies (Consolidation) Act 1908, s 279 which extended the availability of relief to any 'person occupying the position of director.' The original provision in the 1907 Act, which was based upon the Judicial Trustees Act 1896, s 3 (now Trustee Act 1925, s 61), granted the court discretion to relieve a director from liability in negligence. Further subsequent modifications include extending its reach to cover (i) 'default, breach of duty, or breach of trust' (Companies Act 1929, s 372); and (ii) 'an officer of the company or a person employed by the company as an auditor' (Companies Act 1948, s 448). For comparable provisions in other jurisdictions see: Corporations Law, s 1318 (Australia); Companies Act 1955, s 468 (New Zealand).

[3] It does not feature prominently in major textbooks such as P Davies, *Gower's Principles of Modern Company Law* (London: Sweet & Maxwell, 6th ed, 1997) 652; JH Farrar and BM Hannigan, *Farrar's Company Law* (London: Butterworths, 4th edn, 1998) 427–428; and J Birds, AJ Boyle, E Ferran, C Villers, *Boyle & Birds' Company Law* (Bristol: Jordans, 4th edn,

modernisation proposals for company law reform.[4] In its Final Report, section 727 is approached rather obliquely, almost as a footnote to the recommendations concerning minor changes to the permissibility of exclusionary provisions allowed for by section 310 of the Companies Act 1985. Undoubtedly, this may serve to underscore the existence of important inter-connections between court-determined relief from liability and other mechanisms such as a statutory provision governing indemnity, or exoneration provisions (whether contained in a contract between the company and the director or in the company's articles of association).[5] The way in which the matter is dealt with is also explicable by reference to the nature of the provision. As a legal backstop it is more of a peripheral company law concern than the facilitative or regulatory rules that occupy the Steering Group's attention. Somewhat less helpfully, approaching relief indirectly through section 310 tends to obscure precisely what is intended by way of reforming section 727 itself. This caveat aside, it is possible to discern three substantive proposals to change its ambit. The first will re-define who is entitled to be considered for relief by restricting it to directors. This will remove other company officers and auditors from the purview of section 727.[6] Our discussion of the current regime is itself confined to the availability of relief to directors. The Steering Group's second substantive reform addresses the type of breach for which relief may be granted. The aim here is to strip out the detailed statutory language indicating the range of breaches to which relief currently applies and to confine it to those duties laid out in the new statutory statement of directors' duties.[7] If their third change is implemented, one of the two existing pre-conditions to

2000) 545. The provision fares better in practitioner works such as *Gore-Browne on Companies* (Jordans Loose-leaf, 44th ed) at 27.21.5; *British Company Law & Practice* (Bicester, Oxon: CCH Editions Loose-leaf) 32–950; and *Corporate Law Service* (London: Butterworths Loose-leaf) 27.216–27.235. It has been afforded greater attention by MR Pasban, C Campbell and J Birds, 'Section 727 and the Business Judgment Rule: A Comparative Analysis of Company Director's Duties and Liabilities in England and United States Law' (1997) 6 *Journal of Transnational Law and Policy* 201.

[4] *Modern Company Law For A Competitive Economy: Final Report,* DTI/Pub 5552/5k/7/01/NP. URN01/942 and 943, 2001 (hereafter Final Report). This mammoth enterprise culminated in a two volume White Paper *Modernising Company Law* (Cm 5553-I) (hereafter White Paper I) and *Modernising Company Law-Draft Clauses* (Cm 5553-II) (hereafter White Paper II) published on 16 July 2002. Although s 727 is referred to in passing in earlier Law Commissions' Reports (see, *Company Directors: Regulating Conflicts of Interests and Formulating a Statement of Duties*, Law Com No 261, Scot Law Com No 173, 1999 paras 11.40–11.45; and *Fiduciary Duties and Regulatory Rules* Law Com No 236, 1995 para 15.13.), its review was left to the Steering Group. For a commentary on the Steering Group's proposals, see Rt Hon. Lady Justice Arden, DBE, 'Reforming the Companies Acts-The Way Ahead' [2002] JBL 579.

[5] See R Cranston, 'Limiting directors' liability: ratification, exemption and indemnification' [1992] JBL 197. See further, *Gore-Browne on Companies*, n 3 above, para 27.21.

[6] The current provision also applies to liquidators: *Re Home Treat* [1991] BCC 165.

[7] See Final Report, n 4 above, para 6.3 read with para 6.4. For a statement of the proposed statutory duties, see White Paper II, n 4 above, Schedule 2.

the availability of statutory relief, that of reasonableness, will be removed.[8] While the first two changes will narrow its sphere of application, the third is seen as having a potential for making the relieving provision more readily available.

Whatever the value of the minor adjustments to the language of the section, the Steering Group's review is less than explicit in defining the place court-determined relief does and should fulfil within modern company law. It settles upon the conclusion that section 727 is something of a benign mystery, 'little used and still, we believe, of value'.[9] This merits further exploration for a number of reasons. It would be wrong to dismiss this conclusion as merely a perfunctory endorsement of the *status quo* that depends upon conjecture more than tangible evidence. But it does draw attention to the lack of academic analysis of how the provision has fared when it has been judicially considered. Besides, it is essential to approach the specific proposals for section 727 by appreciating the wider aspirations behind the Steering Group's reconceptualisation of company law. Presumably, by recommending the preservation of a mechanism for court-determined relief it tacitly acknowledges its continuing value in assisting with the delivery of productive economic activity; and does so as part of a simplified, accessible, transparent and certain legal framework.[10] A key feature here, which mirrors the state of the modern jurisprudence on the issue,[11] is the move to include a more stringent and objectively based duty of care and skill as part of the codification of directors' duties.[12] This is seen as having consequential implications for the terms and scope of section 727.[13] The work of the Steering Group offers other valuable points of context for evaluating its thinking about the specific direction for sections such as 727. These include an emphasis upon modernising the legal regime so that it prioritises the needs of small enterprises but not to the exclusion of large corporations. Among the reasons proffered for this 'think small first' ethos are: minimising the economic burdens that would attend unnecessary legal changes, and aligning the law to the reasonable expectations of honest business people.[14]

This chapter examines the piecemeal proposals to alter the language of section 727 within the broader context offered by the Steering Group.

[8] See n 91 below.
[9] *Developing the Framework*, DTI/Pub 4754/4k/3/00/NP. URN 00/656, at para 3.77. The provision is not touched on by the White Paper I or II, n 4 above. Presumably it will be dealt with when further draft clauses of the Companies Bill are published.
[10] Final Report, n 4 above, paras 1.16 and 1.21.
[11] See *Re D'Jan of London* [1994] 1 BCLC 561.
[12] See *Developing the Framework*, n 9 above, para 3.66; and White Paper II, n 4 above. See also, S Worthington, 'Reforming Directors' Duties' (2001) 64 MLR 439; and J Birds, 'The Reform of Duties Directors' Duties' in John de Lacy (ed), *The Reform of United Kingdom Company Law* (London: Cavendish Publishing, 2002).
[13] *Ibid*, para 3.77.
[14] *Ibid*, para 153. See also White Paper I, n 4 above, 8–9 and 15.

It does so by briefly tracing its origins. Turning to the judicial contours marking the interpretation of the existing provision, the separate requirements of honesty, reasonableness, and fairness are considered, but not simply as discrete elements governing the availability of directorial relief. The overlap between, and judicial elision of, these statutory terms has, on our analysis, resulted in the section being viewed as an arcane provision. It is also our contention that there are other issues of particular contemporary significance that stand out when considering the case law that has accumulated around section 727. The first, the meaning of the reasonableness requirement, features prominently as part of the current reform agenda. By contrast, our second and third concerns, the applicability of section 727 to director's breaches of a fiduciary character, and its relevance in criminal proceedings, are not dealt with at all by the Steering Group, even though they have featured to a lesser or greater extent in the case law that has been generated throughout the lifetime of the relieving provision.[15] The fourth matter touches a wider ambition. This will be to reflect upon the extent to which section 727 can and should facilitate such cardinal themes that underpin the Steering Group's perception of a modern company law regime, in particular the economic realities that impact upon smaller corporate entities. Court-determined relief that is rooted in delivering a fair outcome can support the need for transparency and accessibility. Fairness is also able to ensure that relief is not granted without proper regard for the legitimate interest of other stakeholders, most notably unsecured creditors.

II. THE ANTECEDENCE OF SECTION 727

The architect of the original relieving provision contained in section 32 of the Companies Act 1907 was the Company Law Amendment Committee 1906,[16] chaired by Lord Loreburn LC.[17] The Committee, whose membership included leading company lawyers of the day such as Palmer and Gore-Browne, was established by the Board of Trade in 1905 to undertake a wide-ranging review of 1862–1900 company legislation. While much of the Report is given over to examining the law governing the issuance of prospectuses, the Committee was mindful that, taken in the round, the reform of company law should be approached with 'great caution, and that

[15] On fiduciary duties and relief see, *Coleman Taymar Ltd v Oakes* [2001] 2 BCLC 749, discussed in the text to n 117 below; as for relief in criminal proceedings, see the discussion beginning in the text associated with n 121.

[16] Cmnd 3052 (London: HMSO, 1906).

[17] Lord Loreburn LC—Sir Robert Reid before his elevation—also chaired the 1895 Select Committee on Trusts Administration which led to the Judicial Trustees Act 1896, s 3 (the predecessor to the Trustee Act 1925, s 61).

while it is desirable by all reasonable means to repress fraud, the utmost care should be taken not unduly to curtail the facilities and advantages under which honest enterprise has for so many years flourished…'.[18] A major anxiety underlying the Committee's deliberations was the diminution in large company registrations between 1900 and 1905. Various reasons were identified by way of explaining this reduction.[19] Of particular significance was the finding that the Companies Act 1900 had introduced stringent provisions relating to the contents of the prospectus and had increased the liabilities of promoters and directors. This, in turn, so it was found, led to a shortage of individuals willing to assume the office of director:

> …many prudent and honest men decline to take the risk of accepting directorships, and rendered the promoter's task in organising his company more difficult.[20]

The Committee recommended a range of reforms designed to inject greater transparency into the companies legislation particularly with respect to the minimum subscription requirement and the liabilities of directors for the contents of the prospectus. While it recognised the need for a legislative response to check and punish dishonesty or gross negligence, the Committee stressed that an appropriate balance had to be struck lest the new legislation be 'turned into an engine of oppression for honest and prudent men.'[21] It therefore went on to recommend a new relieving provision for directors.[22]

Before passing to a contemporary examination of the provision, three initial observations are warranted. First, the general approach of the Loreburn Committee to the question of company law reform resonates with that adopted nearly a century later by the DTI's Steering Group. The Board of Trade's vision in 1906 was that the Committee should 'consider generally by what means Joint Stock enterprise can best be fostered and encouraged, as well as the means by which illegitimate practices may be most effectually repressed.'[23] Nearly a century later, the Steering Group state in its Foreword to the Final Report that company law should provide an effective vehicle for enterprise to flourish freely in a climate of discipline and accountability.[24] A second point relates to the above reference in the Loreburn Committee's Report to the anxiety that without relief the supply of directors might be compromised. There does not seem to be any tangible

[18] N 16 above, para 8.
[19] For example, stamp duty had been increased on the registration of companies by the Finance Act 1899.
[20] N 16 above, para 16(3).
[21] *Ibid*, clause 24.
[22] *Ibid*.
[23] Board of Trade Memorandum, 10 February 1905.
[24] Final Report, n 4, para 9.

contemporary evidence cited to support any such conclusion. It may be that the perception gained impetus from the influence this factor, in terms of the supply of trustees, is thought to have had in leading to the introduction of a relieving provision in section 3 of the Judicial Trustees Act 1896. But it is possible to see why being a director might not have been seen as an entirely appealing prospect in the early years of the twentieth century. Statutes existed which served to supplement a director's common law civil liability;[25] and it was not unknown for the courts to find directors liable to compensate their company even though the breach might be characterised as technical, scrupulously honest, and preceded by legal advice.[26] However, whatever the relevance and prevalence of this anxiety about becoming a director in 1906,[27] it must be highly questionable that it remains a significant rationale for statutory relief today. Broadly speaking, it does not seem credible that the availability of relief for breach of duty influences the choice of business form for the owner-manager. Nor is it likely to figure significantly in decisions about accepting directorships in large companies.

The third observation is of greater import for our purposes and has some relevance to modern developments in which the standard of care expected of directors has been judicially reformulated.[28] Without dissenting from the recommendations, one member of the Loreburn Committee, Sir Edgar Speyer, appended a powerful note to the final Report, headed 'Liability of Directors'. He felt strongly that the liability of directors for negligence in the conduct of business of a company should be increased, particularly in the light of the high volume of liquidations prevailing at that time. Citing a leading German expert of the day,[29] who had stridently criticised the 'intermittent theory' of directors' duties,[30] he stated that 'English law recognises a shadowy liability for negligence and presumes diligence.'[31] It was Sir Edgar Speyer's view that the argument against increased liability in negligence

[25] See the Directors Liability Act 1890, s 3 (liability to compensate for loss occasioned by untrue prospectus statement); and the Companies Act 1900, s 5(2) (liability for irregular allotment of shares).

[26] See *Hirsche v Sims* [1894] AC 654; *Re Faure Electric Accumulator Co* [1889] 40 Ch D 141; *Young v Naval and Military Cooperative Society* [1905] 1 KB 687, cited by Young CJ and Newton J in *Lawson v Mitchell* [1975] VR 579.

[27] For the influence it may have exerted in the reformulation of s 279 of the Companies Act 1908 by the Companies Act 1929, s 372, see the Greene Committee Cmnd 2657, 1926; and C Baxter, 'Demystifying D & O Insurance' (1995) 22 OJLS 537, 542.

[28] See n 80 below, and associated text.

[29] Dr Ernest Schuster of Lincoln's Inn.

[30] See *Re Cardiff Savings Bank: The Marquis of Bute's Case* [1892] 2 Ch 100.

[31] Mr Speyer pointed to the formulation of Lord Lindley MR in *Lagunas Nitrate Co v Lagunas Syndicate* [1899] 2 Ch 392, 435: 'If directors act within their powers, if they act with such care as is reasonably to be expected from them, having regard to their knowledge and experience, and if they act honestly for the benefit of the company they represent, they discharge their … legal duty to the company' (cited with approval by Romer J in *Re City Equitable Fire Insurance Co Ltd* [1925] Ch 407).

to the effect that, 'men will not so easily accept directorships' was no answer. He concluded by calling for a statutory formulation of the standard of care expected of directors together with an express rejection of the notion that their liability for negligence is somehow ring-fenced from the general law.

Notwithstanding the force of his reservations about the state of the common law duties of care and skill, Sir Edgar Speyer nevertheless felt that the new relieving provision was a necessary reform in respect of directors' statutory duties. Yet, it is not entirely easy to distil any single overriding rationale that explicitly prompted the Committee's proposal for a statutory provision offering the possibility of relief to the errant director. The matter is dealt with in clause 24, a single paragraph covering little under a page in the thirty-one page Report, a not dissimilar level of prominence to that which is afforded to section 727 by the Steering Group at the beginning of the twentieth first century. Relief is peripheral to the major thrust of the Loreburn Committee's Report and nothing emerges about the function to be served by the relieving provision outside the realm of statutory duties relating to the company prospectus.

It is the businessman Speyer's commentary on clause 24 that perhaps indirectly offers the best insight into what justified the new regime at its inception. Of these, for present purposes, it is his belief that the law governing directorial liability in negligence was far too lax that is particularly striking. He targets the impossibility of rendering the 'pluralist' and 'ornamental' directors liable to the company for inactivity. Where their companies failed, they were free to move on to their next sinecure. In stark contrast, a sole trader who negligently mismanaged his business would have to suffer the stigma and financial consequences of bankruptcy. These concerns are not of purely historical interest. The types of directors and corporate enterprises may now be different but the controversy about how stringent a director's negligence liability should be has continued unabated in both the law reports and the academic journals.[32] While statutory duties were the Loreburn Committee's primary focus, one reason why section 727 is of modern interest resides in the incidental impact that the shifting judicial approach to the scope of the duty of care has had upon its potential availability to those in breach. It may be argued that the more exacting the duty, the greater the reason for maximising such flexibility as may exist within the statutory language on relief. Such is the tension and potential interplay between duty and relief. For the more congruent they become, the less value there is in the discretionary power to award relief. Ultimately, what is at stake is delivering suitable protection to the company (or its creditors) without

[32] See n 79 below.

making directors pay for honest and reasonable mistakes where that would be unfair.

The statutory outcome of the Loreburn Committee's vision of a relieving provision found in the 1907 Act differed from its precise recommendations. In its original form it favoured a two-tier approach to relief. First, a director (or promoter) in breach of any duty under the prevailing Companies Acts[33] would be eligible for relief if, 'the breach has been occasioned by honest oversight, inadvertence, or error of judgment on his part.' Secondly, were a director to be negligent or commit a breach of trust the Committee proposed a wider relieving power. Such liability might be relieved provided the court was satisfied that he had acted honestly and reasonably. The proposal caused considerable controversy during the House of Commons debates on the ensuing Companies Bill and a considerable body of opinion favoured deleting the provision altogether. However, by way of compromise, the President of the Board of Trade, Lloyd-George, proposed an amendment which deleted only the first tier of the Loreburn provision while keeping the second. Hansard reports him as arguing that:

> Men were made directors of companies for different reasons. A man might be an inventor, for instance and it would be rather hard if, having been made a director because of his expert knowledge, he were held to be responsible for some financial blunder in which he had no part.... So long as each director acted honestly in the special branch of the business allocated to him, he ought to have some protection.... They wanted, after all, to encourage the best men, and they must not frighten them away.... While he did not think they should insert provisions of that kind to frighten desirable men away, he wanted such protection of the public as was given in the Judicial Trustees Act.... It might be that a director of that kind was amply looked after under Subsection (2), taken from the Judicial Trustees Act, and which, no doubt by this time, had stood the test of judicial interpretation.[34]

Section 3(1)(a) of the Judicial Trustees Act 1896,[35] which the Loreburn Committee used as the model for the second tier of their proposed relieving provision, was therefore adopted by the draftsman in framing section 32 of the Companies Act 1907. Although the provision has been subjected to some minor changes,[36] at its heart it remains unaltered. Thus, for relief

[33] The Companies Acts 1862 to 1900.
[34] Hansard, Vol CLXXXI [Fourth Series] HC 893–894, 21 August 1907.
[35] Now the Trustee Act 1925, s 61. For a discussion of the history of this trustee relieving provision, see C Stebbings, *The Private Trustee in Victorian England* (Cambridge: CUP, 2002), ch 6. See also, J Lowry and R Edmunds, 'Excuses' in P Birks and A Pretto (eds), *Breach of Trust* (Oxford: Hart Publishing, 2002), 272–76.
[36] Eg, the Companies Act 1929.

to be available, section 727 of the Companies Act 1985 requires the breach to be honest and reasonable and invests the court with a broad and flexible power to grant total or partial relief from liability if in the circumstances it is fair to do so.

III. THE PRECONDITIONS AND BASIS OF RELIEF

A. Degrees of Overlap

Before tracing the legal development of each of the key requirements, honesty, reasonableness, and fairness separately, this part of the chapter is concerned to demonstrate the existence of overlap between them. In the scheme of section 727 the two pre-conditions are often expressed as elements distinct from fairness, which is the basis upon which the discretion operates.[37] The value of this is in indicating a sequence that is to be followed whereby a determination of the gateway conditions must precede any judicial enquiry into how fair it is to allow full or partial relief. What this judicial approach tends to obscure, however, is that the inter-relationship between honesty, reasonableness and fairness is more subtle, fluid and complex.[38] An illustration of the complexity of the interaction between the requirements can be seen in an important procedural matter. There is authority indicating that it is for the claimant to satisfy the court that he or she is entitled to relief.[39] But, procedurally, the matter is not so straightforward for a number of reasons. First, the onus of proof that applies both to reasonableness and fairness on the one hand differs from that which applies to honesty on the other. Thus in *Re Kirby Coaches Ltd*, Hoffmann J remarked: 'It may be that honesty would be assumed in his favour in the absence of evidence to the contrary, but it is for him to show that he acted reasonably and ought fairly to be excused.'[40] Secondly, because there is no necessity for the section to be pleaded in advance, the defendant cannot be compelled to provide particulars of the facts in support of his assertion that he or she is eligible for relief. Nor will the claimant need to make good this factual deficit once he or she takes the decision to invoke section 727 as part of the legal proceedings. This unsatisfactory position is attributed to

[37] See *National Trustees Co of Australasia v General Finance Co of Australasia* [1905] AC 373, 381; *Re Westlowe Storage and Distribution Ltd (in liquidation)* [2000] BCC 851, 871; *Bairstow v Queens Moat Houses plc* [2000] 1 BCLC 549, 561 (Nelson J), and [2001] 2 BCLC 531, 552, para 63 (Robert Walker LJ); and *Coleman Taymar Ltd v Oakes* n 15 above, 770, para 86.
[38] See *Re Duomatic Ltd.* [1969] 2 Ch 365 where, at 375–376, the failure to take legal advice, commonly regarded within the section's concept of fairness, was taken to affect the reasonableness of the breach.
[39] *National Trustees Co of Australasia v General Finance Co of Australasia*, n 37 above; and *Gamble v Hoffman* (1997) 24 ACSR 606. The application is made by petition: CPR, r 49(1)(i).
[40] [1991] BCC 130, 131.

the inescapable need to follow authority, in the guise of an ill-reported Victorian authority decided on the comparable trustees' relieving provision contained in section 3 of the Judicial Trustees Act 1896, in preference to principle and common sense.[41] It may to some extent now be ameliorated by the Woolf civil justice reforms;[42] but this should not deter the Steering Group from seizing the opportunity to re-consider the question of procedural improvements to the way in which section 727 currently operates.

Seeing honesty and reasonableness as separate requirements also fails to indicate how the courts have sometimes conflated them rather than treated them as discrete considerations. This may occur if it is thought unnecessary to distinguish whether the behaviour complained of is dishonest or unreasonable. For instance, *Bairstow v Queens Moat Houses plc*[43] involved a breach of sections 263 and 264 of the Companies Act 1985. The directors, acting on the company's 1991 accounts that incorrectly showed inflated profits, unlawfully paid dividends which exceeded the available distributable reserves. To the first instance judge, Nelson J, the fact that directors had failed to act honestly and reasonably in preparing the accounts did not preclude him from going on to hold that they were reasonable and honest in authorising some but not all consequential dividend payments.[44] Not surprisingly, the Court of Appeal disagreed. Robert Walker LJ doubted whether there were facts that warranted the need to consider relief and took issue with Nelson J's application of the statutory pre-conditions:

> That sort of schizophrenia is an impossible concept. The judge's reasoning also contains a false syllogism: the former directors paid preference dividends (based on false results); honest and reasonable directors might have paid preference dividends (based on true results); therefore the former directors were acting honestly and reasonably.[45]

Whatever theoretical and other merits there may be in separating out the pre-conditions from the substantive and discretionary element of the relieving jurisdiction, in practice it seems that the two are not mutually exclusive. The concepts and legal tests of reasonableness and fairness share similarities, and factors that contribute to a finding of unreasonableness may also dictate that it is unfair to allow relief. This last point emerges from the treatment of section 727 in *Re Duckwari plc (No 2)*.[46] In essence, Offerventure, a company owned by the defendant director, Cooper, and his wife, sold a disused Methodist church hall to Duckwari for the price Cooper had paid for it, but subject to an agreement that he would have a

[41] *Singlehurst v Tapscott Steampship Co Ltd* [1899] WN 133.
[42] See Part 1 of the Civil Procedure Rules.
[43] [2001] 2 BCLC 531.
[44] [2000] 1 BCLC 549.
[45] N 43 above, 553.
[46] [1998] 2 BCLC 315.

50 per cent share in the profits arising from its development but without any corresponding liability for any loss. The transaction was approved by Duckwari's board, but not by a resolution of the shareholders in general meeting as required by section 320 Companies Act 1985. A collapse in property values forced Duckwari to sell the hall at a loss. In turn the newly appointed Board sought to recoup the loss from Cooper and Offerventure. The Court of Appeal upheld the lower court's refusal to afford Cooper relief.[47] In doing so it did not doubt the honesty of Cooper's breach of section 320. However, in Nourse LJ's view, it was neither reasonable nor fair for Cooper to have a 'one-sided arrangement detrimental to Duckwari' that insulated him (and Offerventure) from bearing any loss on the property development.[48]

On one level it might be thought that this robust conclusion deals rather harshly with Cooper. Presumably his fellow directors, no doubt encouraged by a buoyant property market, willingly entered into the 'profit-and-no-loss' agreement believing that it was financially attractive to the company. As such it was no more than the *quid pro quo* for transferring the hall at the price Cooper had paid for it. Admittedly, this might be precisely the type of circumstance which section 320 is designed to address namely, to protect the shareholders interests. Had they sanctioned the deal, Cooper would have escaped liability altogether. But, of course, the shareholders did not have that opportunity. On that basis the unavailability of section 727 can serve to underscore the importance of disclosure and ratification mechanisms.

B. Honesty

Although the inter-connections between honesty, reasonableness and fairness is a central thread of this chapter, we will nevertheless consider in turn what the case law tells us about each of them. There is a dearth of judicial discussion about what, as a matter of law, 'honesty' means for the purposes of section 727. This may reflect the practical reality that this is regarded as the less troublesome of the two pre-conditions, its presence or absence on the facts being self-evident, or presumed unless there is evidence to the contrary.[49] Although the extent and materiality of the consideration in the case law varies in its degree of intensity and directness, determining either extremity of honesty or dishonesty has seldom proved contentious. Detailed judicial exposition on how the directors' behaviour is to be considered honest is rare; and when the issue does occasionally surface in judicial

[47] [1997] 2 BCLC 729.
[48] N 46 above, 326.
[49] N 40 above.

deliberations it does so generally by way of *obiter dictum*. Thus in *Re Welfab Engineers Ltd*,[50] Hoffmann J held that the directors' failure to realise the best price for the company's principal asset, its freehold premises, was an honest attempt to save the business and the jobs of the employees. He therefore dismissed the liquidator's misfeasance summons brought under section 212 of the Insolvency Act 1986. Although this made it unnecessary for him to consider section 727, he nevertheless took the opportunity to signal his willingness to relieve the directors had it been necessary to do so. Doubtless it is safe to conjecture that Hoffmann J's inclination to see the behaviour as honest coloured his brief excursion into the potential application of the relieving provision. At the other end of the spectrum, *Dorchester Finance Co Ltd v Stebbing* offers a direct illustration of patent dishonesty.[51] Even though the director, Stebbing, declined to give oral testimony, Foster J had little difficulty in determining that the defendant's knowing and reckless misapplication of £400,000 worth of company assets amounted to a dishonest breach of the duty of care and skill which did not, therefore, qualify him for relief under section 727.[52]

As commented above, notwithstanding the prominence that the honesty requirement receives in the drafting of the provision, there has been little judicial cause either to analyse it precisely or settle conclusively the degree of subjectivity or objectivity with which it is to be measured. What amounts to dishonesty is touched upon by way of *obiter dictum* in *Re J Franklin & Son Ltd,*[53] where relief was sought when the annual general meeting's resolution to pay Mrs Franklin remuneration as director of £8 per week was impugned by the liquidator because one member voting in its favour had not been technically registered as such. Without doubting the honesty of those directors who had wrongly authorised the payment, Crossman J observed:

> I read the word 'dishonestly' in the section as meaning without any direct motive. If they had done this in order to assist Mrs Franklin recklessly, but without considering the interests of the company at all, I think that they might well be said to be dishonest, but I do not find that the evidence goes to that.[54]

Leaving aside the way in which the judge reads the statutory wording as if it refers to dishonesty rather than honesty, the language conveys the sense in which any breach that is accompanied by the highest degree of *mens rea* is outwith the precondition of honesty. This appears to rely upon notions

[50] [1990] BCLC 833.
[51] [1989] BCLC 498. See also, *Selangor United Rubber Estates Ltd v Cradock (No 3)* [1968] 1 WLR 1555, 1660.
[52] *Ibid,* 505–06.
[53] [1937] 4 All ER 43.
[54] *Ibid*, 47.

most familiar to criminal law. It does not address the fundamental issues of whether this is entirely apposite for a relieving provision that is largely concerned with breaches of civil liability and whether or not any assessment of honesty is to be made in a more or less objective fashion. Admittedly, on the latter point, the Court of Appeal, in *Bairstow v Queens Moat Houses plc,*[55] suggests that the trial judge's finding that honesty is to be assessed from an essentially subjective standpoint,[56] is open to question. Without reaching a definite conclusion on this point, Robert Walker LJ states his preference for an objective test to determine a director's honesty,[57] by citing the following *obiter dictum* from Lord Nicholls' speech delivered in *Royal Brunei Airlines Sdn Bhd v Tan*,[58] a leading authority on the approach in trust law cases dealing with accessory liability for breach of trust:

> The standard of what constitutes honest conduct is not subjective. Honesty is not an optional scale, with higher or lower values according to the moral standards of each individual. If a person knowingly appropriates another's property, he will not escape a finding of dishonesty simply because he sees nothing wrong in such behaviour.

When the point is expressed in such vivid and persuasive language it is easy to understand why there should be judicial reluctance to countenance allowing a defendant's own perceptions to influence the judicial determination of honesty for the purposes of relief. However, it is worth sounding two points of caution. The first concerns ambiguity inherent in the taxonomy of subjective and objective. It should not be forgotten that in practice these are not always clearly delineated but rather coexist as a more subtle amalgam that is harnessed to determine honesty or, for that matter, reasonableness. There is then ample scope for misunderstanding because it is not always clear-cut whether the tests propounded by different judges (or commentators) coincide. They may have in mind varying degrees or shades of subjectivity and objectivity. Three possible standards have been judicially articulated. The purest form of subjectivity finds expression in what is termed the 'Robin Hood test', under which there is no dishonesty where the director genuinely believes his or her behaviour is morally justifiable even though it flouts the standards of reasonable honest people. Although taken out of context, the passage cited above from Lord Nicholls' speech appears to amount to a rejection of a purely subjective assessment of honesty. But although Nelson J in *Bairstow* seems to speak of honesty being determined in an essentially subjective way, he does so by reference to

[55] N 43 above.
[56] N 44 above, 572, citing the approach of Knox J *Re Produce Marketing Consortium Ltd* [1989] BCLC 513, at 518, which also found favour more recently in *Coleman Taymar Ltd v Oakes,* above n 15, 770.
[57] N 43 above, 550.
[58] [1995] 2 AC 378, 389.

cases on reasonableness, in which the test has both objective and subjective components.[59] Robert Walker LJ's tentative rebuttal of the essentially subjective test may therefore have been misplaced because it obscures the objective element. In effect the difference may be more linguistic than substantive.

A second related reason for caution in construing the word honesty in section 727 by reference to this particular quotation from *Tan* is because the House of Lords in *Twinsectra Ltd v Yardley* has since subjected Lord Nicholls speech to extensive scrutiny.[60] In the interpretation favoured by the majority, the determination of dishonesty is neither a purely objective nor subjective investigation, but depends upon what Lord Hutton terms 'the combined test'.[61] Drawing upon criminal law authority,[62] this asks first, if the defendant's conduct was objectively dishonest by reference to the ordinary standards of reasonable people. If the answer is in the affirmative, the second part of the combined test is applied by determining whether there is evidence that the defendant realised that he or she had behaved dishonestly in the circumstances. As Lord Hoffmann emphasises, the principles require more than simply showing dishonest conduct, '[t]hey require a dishonest state of mind, that is to say, consciousness that one is transgressing ordinary standards of behaviour'.[63]

If this is to be applied to section 727, it has the novel merit of identifying the appropriate standard in the most explicit and lucid way to date. This is not to claim that the test so expressed is necessarily new or different from that which has been in the mind of the relatively few judges who have hitherto addressed the matter. Drawing a conclusion is not easy because honesty has rarely been the determining factor in the denial of relief. Robert Walker LJ's rejection in *Bairstow* of a purely subjective test seems appropriate in the context of determining eligibility for the exercise of relief under section 727. Assessing whether a director has acted honestly by reference to his or her own genuinely held, albeit objectively perverse, conception of morality (the Robin Hood test), would undermine the operation of the relieving provision. However, there is nothing in the reasoning in a case such as *Franklin* that is at odds with the combined test. It is likely and proper that the defendants would have been regarded as having been honest by reference to the ordinary person's conception and, though it would therefore not be strictly material to proceed to the subjective element of the combined test, the defendants would probably have been able to demonstrate that they believed that their actions were honest. By contrast,

[59] See *Re Produce Marketing Consortium Ltd,* n 56 above.
[60] [2002] 2 WLR 802.
[61] *Ibid*, 808, para 27, a phrase adopted by Lord Steyn (805, para 7). The approach is also endorsed explicitly by Lord Hoffmann (paras 20–23) and indirectly by Lord Slynn (804, paras 3 and 5).
[62] *R v Ghosh* [1982] QB 1053.
[63] N 60 above, 807, para 20.

it seems hardly credible that the misuse of company funds by the director Stebbings in *Dorchester Finance* could be regarded either as objectively honest or subjectively so. Moreover, it is suggested that the combined test may not, in substance at least, differ much from Nelson J's essentially subjective test.

If, as the Steering Group proposes, reasonableness is deleted, leaving honesty as the only gateway condition in a reformed section 727, there is added value in having an authoritative elaboration of what the single remaining pre-requisite means. And, given the scope for taxonomical ambiguity, it might be even more valuable if any new statutory formulation expressly identifies how honesty is to be defined for the purposes of relief. However, this raises the crucial substantive question of whether the combined test is the most appropriate one, or whether the rationale of section 727 might be better served by a more objective formulation such as that preferred in Lord Millett's vigorous and provocative dissenting speech in *Twinsectra*. This view, itself claiming to be reflected in Lord Nicholls' speech in *Tan*, eschews all reference to whether or not the defendant realises his or her conduct is dishonest.[64] The test is not wholly objective, for the defendant must have knowledge of the facts which, objectively, make the conduct dishonest. So in that restricted sense the test is subjective, relating '… to the defendant's knowledge, experience and attributes.'[65] As such it accords with the view that civil liability should turn on an equity lawyer's perspective of dishonesty rather than one more relevant to criminal law, because '… equity looks to a man's conduct, not to his state of mind.'[66] There are attractive consequences attaching to this conduct-driven test. It is probably easier to apply in practice and, as a matter of principle, the influence upon this test of the defendant's own personal standards of honesty is minimised. It is also a more exacting standard. However, on balance it is contended that the combined test is preferable for the purposes of section 727. The provision confers upon the court a general relieving jurisdiction. It does not contain a specific form of equitable liability. The defendant will have been found liable, often by reference to a test, such as that of the duty of care and skill, which is objectively dominated. Applying a too objectively orientated test in respect of any pre-condition of relief may therefore render the value of the section nugatory. Indeed, in assessing eligibility for relief, looking to the defendant's state of mind should be a major second strand to an objectively based requirement of honesty. It is our contention that this is best catered for by the combined test.

[64] *Ibid*, 836, para 121.
[65] *Ibid*, para 122.
[66] *Ibid*, 837, para 123.

C. Reasonableness: An Objective or Subjective Determination?

As a matter of policy and judicial opinion, it seems to be settled that for section 727 a director's honesty will be determined by a combined test that is a predominantly objective standard. There is, perhaps surprisingly, less certainty about how the comparable issues are to be addressed in respect of reasonableness. Unlike honesty, this issue has figured much more prominently in the section 727 case law on reasonableness. For some considerable time the authorities inclined in favour of a substantially subjective test of reasonableness, but there appears to be something of a sea-change underway.

The judicial tendency to dilute the objective character of the concept of reasonableness in determining the availability of relief is most often attributed to the influential decision in *Re D'Jan of London Ltd*.[67] Hoffmann LJ, sitting in the Chancery Division, took a broad subjective view of the conduct in question for the purposes of determining whether the director had acted reasonably. A straightforward proposal form for property insurance contained numerous factual misrepresentations and the insurers therefore repudiated liability when the company claimed for fire damage. Mr D'Jan, the controlling director, who together with his wife held the entirety of the company's shares, had signed the proposal without reading it. This prompted, by way of a misfeasance summons, the question of D'Jan's liability for breach of the common law duty of care and skill. Hoffmann LJ assimilates the common law standard of care with the two-limb test contained in section 214(4) of the Insolvency Act 1986. On this basis D'Jan's liability was clear cut. The more taxing question was whether or not D'Jan should be relieved from liability. Turning to section 727 the learned judge observes:

> It may seem odd that a person found to have been guilty of negligence, which involves failing to take reasonable care, can ever satisfy the court that he acted reasonably. Nevertheless, the section clearly contemplates that he may do so and it follows that conduct may be reasonable for the purposes of section 727 despite amounting to lack of reasonable care at common law.[68]

This was not the first time that the oddity had been identified. Looking at the parallel trust legislation, Chancery judges had been moved to speculate that the conundrum inherent in regarding negligence as being reasonable

[67] N 11 above.
[68] *Ibid*, 564. This conundrum has also been identified, but left unresolved by antipodean courts: see *Fletcher v National Mutual Life Nominees Ltd* [1990] 3 NZLR 641, at 692–93; *Dimond Manufacturing Co Ltd v Hamilton* [1969] NZLR 609; and *Pacific Acceptance Corporation Ltd v Forsyth* (1970) 92 WN 29, NSW.

was the product of defective drafting.[69] It is suggested that it is implicit, rather than explicit, within the language of Hoffmann LJ's solution that reasonableness admits of subjective connotations. This does not deny an attractive pragmatism and neatness in Hoffmann LJ's approach that can be sustained in the genesis of section 727. And although Hoffmann LJ seems to reason more by reference to principle than precedent, it is possible to read earlier cases on the company law relieving provision as adding a subjective gloss to its pre-requisite of reasonableness. For instance, in *Re Gilt Edge Safety Glass Ltd,*[70] a reduction in the company's share capital had the unforeseen consequence of depriving two directors of the minimum permissible share qualification as required by the articles. They nevertheless continued to serve for approximately two years. When Triplex Safety Glass Ltd subsequently acquired the company it sought to recover the sums wrongly drawn by way of directors' remuneration throughout the two year period. The two directors successfully claimed total relief from any prospective liability within the jurisdiction provided by the relieving provision, then contained in section 372(2) of the Companies Act 1929, notwithstanding the opposition of the members of the company. Without addressing the matter as explicitly as *Re D'Jan* does, Crossman J appears to have a similar understanding that though the breach is objectively negligent, it could still be reasonable to relieve the wrongdoers:

> I think that in a sense it may be that they were negligent, and no doubt, without knowing it, they committed a breach of duty in continuing to act as directors when they were no longer qualified. But this was a purely technical defect which would if it had been realised, have been put right at once … .[71]

Subsequent cases followed Hoffmann LJ's lead in *Re D'Jan* without further question.[72] Comparatively recently, subjectivity significantly and peculiarly coloured the interpretation of section 727 of the trial judge in *Re Simmon Box (Diamonds) Ltd.*[73] This misfeasance action under section 212 of the Insolvency Act 1986 included an allegation that the 19 year old 'sleeping'

[69] *Perrins v Bellamy* [1898] 2 Ch 521, 528–29.
[70] [1940] 1 Ch 495.
[71] *Ibid,* 503. See also the similar, earlier case, *Re Barry and Staines Linoleum Ltd* [1934] Ch 227, in which Maugham J observes (233–34), '… the petitioner has in my opinion acted honestly and reasonably, notwithstanding that there was a certain negligence in his not ascertaining that the articles of association required him to obtain his qualification … shares within two months after his appointment as a director.' It is noteworthy that relief was granted from the criminal penalties imposed by the Act for failure to obtain the necessary qualification shares.
[72] That the test for s 727 is subjective has indirectly surfaced in relation to the wrongful trading provision, s 214(1) of the Insolvency Act 1986. See *Re Produce Marketing Consortium Ltd*, n 56 above. In *Re Brian D Pierson (Contractors) Ltd* [1999] BCC 26, Hazel Williamson QC, sitting as a Deputy High Court Judge, applied *Re D'Jan*, n 11 above, observing (at 48) that: '…'reasonableness' for the purpose of s 727 must be meant to be capable of being satisfied by something less than compliance with the common law standard of care in negligence.'
[73] [2000] BCC 275.

director, Andrew, had been negligent in delegating to his fellow director (and father) management tasks including the insurance of valuable gemstones which were lost by the father on a cross channel ferry. Peter Smith QC, sitting as a Deputy Judge of the Chancery Division, drew extensively upon Hoffmann LJ's subjective formulation of the concept of reasonableness. Although the judge did not explicitly adopt the test, it seems incontrovertibly to influence the grant of partial relief to Andrew. Viewed objectively, Andrew had not behaved as a reasonable director. However, on Peter Smith QC's reasoning, he deserved some relief from liability because he was a student living overseas who did not participate in the management of the company beyond occasionally signing cheques drawn on the company's bank account. The judge took the view that the company's losses were not caused by any direct fault on his part, but arose from the conduct of his father in whom, it was found, 'he reposed too much trust.'[74]

Whether reasonableness in section 727 should or should not be read as subjectively as *Re D'Jan* suggests has recently been judicially challenged. The first dissentient voice is heard in *Bairstow*, where Robert Walker LJ refuted the notion that reasonableness is capable of being considered from an essentially subjective point of view.[75] This change of tack was followed in the first instance decision in *Coleman Taymar Ltd v Oakes,*[76] in which Judge Robert Reid QC, sitting as a judge of the High Court, observed, 'I do not see how the reasonableness requirement can be a subjective requirement. Any reasonableness test must by its very nature be objective.'[77] We have already indicated the need for taxonomical caution when interpreting such references to objective and subjective tests. In any event, although it may be too early to claim that these judicial statements are tantamount to a *volte-face*, they are definite portents of a more critical approach being taken towards the determination of this fundamental requirement of the relieving provision.

The unsettled state of the case law reflects the silence of the statute in defining what meaning it intended for the pre-condition of reasonableness, both as to where the standard should be pitched and the extent to which its determination should be more or less objective than subjective. Over the years, judicial preferences have emerged in the form of terse observations

[74] In the event, relief under s 727 became irrelevant because the Court of Appeal decided that the judge had strayed beyond the pleadings, and accordingly found that Andrew was not liable at all: *Cohen v Selby* [2001] 1 BCLC 176. Interestingly the first instance judge saw potential in the relieving power to be a mechanism 'to achieve a fairness as between wrongdoers': *Ibid*, 288. Similarly, the comparable Australian provision in the Corporations Law, s 1318 has been canvassed as an alternative to contributory negligence where such a claim is not recognised: *AWA Ltd v Daniels t/as Deloitte Haskins & Sells* (1992) 7 ACSR 759; cf the effect of the claimant's conduct in assessing the fairness of relief in *Coleman Taymar v Oakes* n 15 above, 771–72, para 100.

[75] N 43 above, 550, para 58.

[76] N 15 above.

[77] *Ibid*, 770, para 85.

without much in the way of supporting rationale. While this may be inevitable, in part, because relief is commonly less prominent than the question of liability, this leaves the matter tantalisingly unresolved. At first sight it may appear obvious that reasonableness as a pre-requisite to eligibility for relief should turn more on objective than subjective considerations. The point may seem to gain force where the breach itself involves a determination that the director has behaved unreasonably, and such unreasonableness falls to be assessed in a predominantly or exclusively objective fashion. Thus in *Re D'Jan* the breach concerned the common law duty of care and skill rather than some statutory duty under the Companies Act. Since *Re City Equitable Fire Insurance Co*,[78] the proper scope of the duty, in particular its subjective/objective credentials, have proved a fertile source of judicial and academic discussion.[79] Hoffmann LJ, however, sought to settle the matter, assimilating the common law with the heavily objectively laden language of the wrongful trading provision, section 214 of the Insolvency Act 1986.[80] If, as a consequence, a synergistic approach were then to prevail in defining section 727's use of the term reasonableness, it would effectively preclude all directors in breach of the common law duty from being eligible for consideration of total or partial relief. It is, presumably, this that inspired Hoffmann LJ to suggest that the section clearly contemplates that relief may involve a more subjective idea of reasonableness than applies in the determination of reasonable care at common law. Although inevitably a matter of conjecture, this approach would seem to respond to the anxiety

[78] N 31 above.

[79] See *Daniels v Anderson* (1995) 16 ACSR 607, NSWCA. For academic commentary see, for example, C Riley, 'The Company Director's Duty of Care and Skill: The Case for an Onerous but Subjective Standard' (1999) 62 MLR 697; V Finch, 'Company Directors: Who Cares about Skill and Care' (1992) 55 MLR 179; B S Butcher, *Directors' Duties: A New Millennium, A New Approach?* (Deventer: Kluwer, 2000) 259–261; S. Worthington, 'The Duty to Monitor: A Modern View of the Director's Duty of Care' in F Patfield (ed), *Perspectives on Company Law: 2* (London: Kluwer Law International, 1997); A Walters, 'Directors' Duties: the Impact of the Company Directors Disqualification Act 1986' [2000] *Co Law* 110; L Griggs and J Lowry, 'Finding the Optimum Balance for the Duty of Care Owed by Non-Executive Directors' in F Patfield (ed), *Perspectives on Company Law: 2* (London: Kluwer Law International, 1997); and J Birds, n 12 above.

[80] Indeed, prior to the decision in *Re D'Jan,* Hoffmann J had stated that he was 'willing to assume' that the wrongful trading provision provided an accurate statement of the director's common law duty of care and skill: *Norman v Theodore Goddard (a firm)* [1991] BCLC 1028, 1030–31. The reasoning proceeds as follows: by way of defence to a wrongful trading claim, s 214(3) provides that the court will not hold a director liable if, once he found himself in a position where he knew or ought to have known that the company was going into insolvent liquidation he took every step with a view to minimising the potential loss to the company's creditors. The facts which a director ought to know or ascertain for the purposes of s 214(3) are determined predominantly by way of objective assessment (s 214(4) refers to 'a reasonably diligent person' as the principal criteria). The judicial shift towards objective assessment of directors' behaviour is also apparent in relation to cases brought under the Company Directors Disqualification Act 1986: see, *Re Barings plc (No 5)* [1999] 1 BCLC 433; [2000] 1 BCLC 523 (Chancery Division & Court of Appeal); *Re Landhurst Leasing plc* [1999] 1 BCLC 286. This development has not been confined to English law; see *Daniels v Anderson* (1995), n 79 above.

expressed by at least one member of the Loreburn Committee, Sir Edgar Speyer, who appears to view the relieving provision as a necessary trade-off for increasing the standard of care required of directors.[81] However, such a stratagem results in the duty of care and skill influencing the interpretation of section 727 for all other kinds of directorial breaches of duty that do not necessarily entail issues of reasonableness. Alternatively it might be open to develop the concept of reasonableness within section 727 so that it is interpreted differently according to whether liability is for breach of the common law duty of care and skill or some other statutory duty. However, this might prove to be as practically cumbersome as it is theoretically unappealing.

D. Section 727 and the Shadow of the Wrongful Trading Provision

In reasoning about the scope of liability at common law *Re D'Jan* draws creatively upon a significant twentieth century statutory provision, section 214 of the Insolvency Act 1986. However, this is not the only point of contact between that statute and the general discretion to grant relief from liability. The dispute as to whether reasonableness is to be assessed subjectively or objectively for the purposes of section 727 has assumed a wider significance in the development of the reach of the modern regime governing wrongful trading. Three years after the introduction of the new claim for wrongful trading had been introduced, Knox J in *Re Produce Marketing Consortium Ltd*,[82] was faced with the legal issue of whether or not the different legislative provisions operated together.[83] The task was not made easy because neither statute explicitly addressed how they were meant to relate one to another. Having considered the matter with some care, the learned judge decided in effect that it could not have been Parliament's intention that the two sections were compatible.[84] For present purposes, one significant reason given in support of this conclusion is the disparity evident in the predominantly objective notion of what is reasonable within

[81] N 31 above and associated text.
[82] N 56 above.
[83] Similar questions were considered by the Court of Appeal concerning the availability of s 727 in respect of proceedings under the Betting and Gaming Duties Act 1972, s 2, in *Customs & Excise Commissioners v Hedon Alpha Ltd* [1981] QB 818.
[84] Although in *Re DKG Contractors Ltd* [1990] BCC 903, Judge J Weeks QC, sitting as a deputy judge of the High Court, did not rule out the availability of s 727 relief to liability under the Insolvency Act 1986, s 214. In comparison, in *Re Duckwari plc (No 2)*, n 46 above, the Court of Appeal, without deciding the point, tentatively endorses the first instance judge's view (reported at [1997] BCLC 729) that there is nothing in the Companies Act 1985, s 322(5) and (6) concerning liability for substantial property transactions that excludes the operation of s 727. There is authority that the Australian relieving provision (Corporations Law, s 1318) can be relied upon by a director in breach of the duty to prevent insolvent trading by his or her company: *Kenna & Brown Pty Ltd v Kenna* (1999) 32 ACSR 430.

section 214 and the subjectivity with which the courts have imbued section 727's notion of reasonableness. In short it was decided that there is no place for subjectivity as a means of offering relief to a negligent director where the liability arises from his or her company's insolvency. Although the detail of the judicial reasoning in this decision has been subjected to some penetrating and persuasive criticism, the outcome has much to commend it.[85] Overall, it seems to be a logical exception to the generosity inherent in *Re D' Jan's* attitude towards section 727 and directorial negligence where the company is solvent.[86] It certainly works from the premise that subjectivity has a role in assessing reasonableness for the purposes in section 727. It might therefore be argued that, in the light of the most recent Court of Appeal *obiter dictum* in *Bairstow* preferring a purely objective meaning for reasonableness, the rationale for denying section 727 relief to defendants in wrongful trading claims has evaporated. This may not much matter in practice because it is arguable that many of those who are found liable for wrongful trading are unlikely to be able to present a strong case for relief.[87] However, as matter of policy, allowing relief under section 727 would surely be an unfortunate denouement, out of keeping both with the gravity of the wrongful trading provision and the current government proposal to keep the company and insolvency law regimes distinct.[88] It would also run counter to the decision by the framers of section 214 to reject the Cork Committee's recommendation[89] for a specific relieving power where the wrongful trading had been honest and the court decided that the errant director ought in all the circumstances to be relieved from liability.[90]

E. Whither Honesty and Reasonableness?

Regardless of whether the context is wrongful trading or some other company law default, the function and meaning of reasonableness in section 727 requires clarification. The Steering Group would drop the pre-condition entirely.[91] This proposal does not seem to spring from the state of flux

[85] See JR Bradgate and G Howells, 'No Excuse for Wrongful Trading' [1990] JBL 249.
[86] Generosity that Hoffmann LJ acknowledges explicitly: n 11 above, 564.
[87] See JR Bradgate and G Howells, n 85 above, 252; MR Pasban *et al*, n 3 above, 205; and S Griffin, *Personal Liability and Disqualification of Company Directors* (Oxford: Hart Publishing, 1999) 77.
[88] See White Paper I, n 4 above, paras 3.12–3.13.
[89] Sir Kenneth Cork CBE, *Insolvency Law & Practice: Report of the Review Committee*, Cmnd 8558 (London, HMSO, 1982), paras 1793–94.
[90] The relevant White Paper, *A Revised Framework for Insolvency Law* Cmnd 9175, para 52 noted, 'There will be no provisions for anticipatory relief and the Government considers that, as the Court will have a general discretion to determine the extent of personal liability, no additional form of relief is necessary.'
[91] *Developing the Framework*, n 9 above. This proposal proved relatively non-contentious with the majority of the forty or so respondents to Question 3.15 in that consultation document.

evident in the case law on how far the concept should be objectively rather than subjectively determined. On the contrary, it consciously accepts the subjective interpretation espoused in *Re D'Jan* and aspires to put beyond doubt that relief should be possible where the cause of action concerns negligence. Deleting the requirement of reasonableness will have the additional virtue of curtailing the judicial debate. However, a bolder approach would be to go further by also deleting the requirement of honesty thus obviating any further controversy about the appropriate level of objectivity to be applied in its determination. This would have the advantage of moulding relief around a more open-textured jurisdiction that gives the court free rein to weigh up the equity or fairness of allowing relief. This sounds a more radical step than it is in substance. Dishonest and incompetent directors are likely to struggle in sustaining the claim that it is fair to relieve them from liability. To this extent it might not advance either cause in the subjective/objective debate in respect of either honesty or reasonableness. It will nonetheless circumvent some of these difficulties, and lend a greater measure of transparency to the operation of the jurisdiction than has perhaps been displayed hitherto. For, it is our contention that a common feature of the judicial discussion of section 727 has, contrary to the conventional learning, failed to keep the pre-conditions, especially that of reasonableness, separate from fairness.[92] This elision may offer a more convincing explanation for the outcome in cases such as *Re D'Jan*. It also provides a workable and theoretically valuable basis for the continuing jurisdiction, whereby fairness may be informed, but not hamstrung, by factors such as how honest and reasonable the breach is.

F. Fairness

On balance, any suggestion in the case law that the three requirements contained in section 727 have been intermingled does not emerge clearly from the more prosaic learning that has accumulated about what fairness means for the purposes of the section. On a conventional reading the breadth of

Amongst the minority viewpoint some harboured reservations about deleting reasonableness (PricewaterhouseCoopers), whilst others were totally opposed to its deletion on the basis that this would allow the (presumably objectively) incompetent director to become eligible for relief (TUC and Institute of Chartered Accountants of Scotland): Responses to Developing the Framework (available at http://www.dti.gov.uk/cld/reviews/urn00656.htm). Interestingly, this would in effect mirror the Cork Committee's lost proposal for a specific relieving provision for wrongful trading and replicate the language of the Australian relieving provision in the Corporations Law, s 1318.

[92] See *Re Simmon Box (Diamonds) Ltd,* n 73 above; although this is not always the case see, for example, *Re Westlowe Storage Distribution Ltd,* n 37 above, where Hart J declined to consider exercising the relieving jurisdiction because the defendant's failure to oversee proper accounting systems and his making of an informal and undisclosed loan were unreasonable.

this discretion has proved relatively uncontroversial. The received wisdom draws on the judicial approach to section 61 of the Trustee Act 1925 that acknowledges the impossibility of distilling any firm guidelines to inform the operation of the discretion.[93] Consequently, it is not possible to define clear categories of breach that will always be excused. Nor is it safe to speculate on the circumstances surrounding either the conduct giving rise to the breach or the particular defendant that will necessarily trigger its exercise. This is no doubt desirable lest the standard of care of directors is to avoid being aligned with *ad hoc* successful appeals to relief. Equally, it does little to illuminate the underlying rationale that justifies the existence or scope of the provision; and on a more practical level it does not assist those who seek professional guidance on the availability of relief in their case.

A closer look at the jurisprudence generated, both by section 727 since its introduction in 1907 and the analogous trustee provision, does afford some insight into the factors that will impact upon judges when deciding whether or not to exercise the statutory discretion. With respect to section 61 of the Trustee Act 1925 it seems that there is a consensus of judicial opinion that the court should be loath to excuse the liability of remunerated trustees unless they have taken all reasonable steps to remedy their breach of trust.[94] Although the issue has not directly arisen in the company law context, there is no reason to suppose that a similar approach will not be taken towards a paid director receiving ample recompense for the risks he or she takes. But that is not to say that a remunerated director will never be entitled to relief.[95] A further significant factor which emerges from the section 727 case law,[96] is that a director who has sought legal advice before proceeding with what amounts to a technical breach of duty is likely to be viewed favourably.[97] In *Re Claridge's Patent Asphalte Co Ltd,*[98] the parent company's directors had expended company money in promoting a subsidiary. The court assumed that this was *ultra vires* the objects of the parent company, although the directors had acted on the advice of the company's solicitors and counsel. The liquidator therefore brought a misfeasance claim

[93] *Re Kay* [1897] 2 Ch 518, 524; and *Re Turner* [1897] 1 Ch 536, 542.

[94] *National Trustees Company of Australasia Ltd v General Finance Company of Australasia Ltd*, n 37 above, 381. See also, *Re Rosenthal* [1972] 1 WLR 1273. This may explain why the liquidator was refused relief in *Re Windsor Steam Coal (1901) Ltd* [1929] 1 Ch 151.

[95] In context of s 61 the Court of Appeal in *Re Pauling's Settlement Trusts* [1964] Ch 303 granted partial relief to paid trustees who were bankers. Relief was granted to a professional trustee company in *Re Te Huango* [1993] 3 NZLR 77.

[96] S 61 decisions appear more equivocal on this point in so far as the judges have indicated that acting on legal advice is not a 'passport to relief': *Marsden v Regan* [1954] 1 WLR 423, 434–435; and see also, *Baden, Delvaux and Lecuit v Société Général pour Favoriser etc* [1993] 1 WLR 509, 609. However, in *Re Evans (dec)* [1999] 2 All ER 777 it was held that the defendant trustee had been reasonable in seeking and relying upon her solicitor's legal advice and ought, therefore, to be partially relieved from liability for breach of trust.

[97] Conversely, failing to take advice may result in relief being denied: see *Re Duomatic Ltd*, n 38 above; and *Coleman Taymar Ltd v Oakes* n 15 above, 768, para 73.

[98] [1921] 1 Ch 543.

against them for breach of trust arguing that the relieving provision had no application to *ultra vires* transactions. Astbury J, rejecting the liquidator's contention on the basis that the wording of the section was wide enough to cover all breaches of trust, granted relief observing:

> This transaction, if one can differentiate one breach of trust from another, was as little harmful or improper as one can very well imagine. The then directors did their utmost for the benefit of their *cestui que trust* company; they took the best advice; they acted in the way complained of openly for no benefit to themselves, but for the benefit of their *cestui que trust*; if it were not for a mere technicality, no objection could be taken to the transaction at all; and apart from the effect of the war, no such objection would probably have been thought of.

The sympathy of the court was patently with the defendants. Apart from the reference to the unforeseeable and cataclysmic world events (the Great War), there are at least two other striking features of the decision to allow total relief. They had acted reasonably in taking professional advice. Their prudence was therefore not merely a precondition to relief—it was material to the assessment of the fairness of affording it. What else is noticeable is the technicality of their breach. Their enterprise had been open, sanctioned by the shareholders, had not of itself proved harmful to the parent company, nor had it been motivated by a desire to secure any personal benefit. Astbury J's reasoning highlights that the determination of relief is necessarily a fact intensive exercise. In this respect the decision does not stand alone. Characteristics such as the technicality of the breach, the absence of loss and acting on expert advice figure prominently in allowing relief from prospective liability in *Re Gilt Edge Safety Glass Ltd*.[99]

Re D'Jan goes further by showing how extrinsic considerations, such as the so-called economic realities, may indicate why fairness dictates that the discretion should be exercised.[100] Two factors were critical to Hoffmann LJ's reasoning. First, Mr D'Jan held 99 per cent of the shares in the company

[99] N 70 above. See also, *Coleman Taymar Ltd v Oakes,* n 15 above, 771–72; and *Clydebank Football Club Ltd v Steedman* 2002 SLT 109. One case that is perhaps difficult to square with the emphasis in granting relief on the basis of the breach being a mere technicality is *Re J Franklin & Son Ltd*, n 53 above (here the judge was the same as in *Re Gilt Edge*). The reasoning for refusing relief for any of the three defendants is not spelt out in the judgment. Acting on an accountant's advice may assist a director's claim: see *Miller v Belmil Products Ltd* [1976] 1 NZLR 311, 320. But the court may be reluctant to conclude in his or her favour where a personal gain has been made: see *Re Duckwari plc (No 2),* n 46 above; see also Tipping J in *Hilton International Ltd v Hilton* [1989] 1 NZLR 442, 478: 'It is one thing for directors to be excused for an honest and reasonable blunder which causes the company loss. It is quite another matter if the same blunder has not only cost the company money but also put that money into the directors' pockets.'

[100] It seems that economic realities might even encompass socio-economic circumstances such as recession and the effects of regional unemployment: See *Re Welfab Engineers Ltd*, n 50 above, 838.

and had proved as an unsecured creditor of the company in the sum of £102,913. Secondly, his breach of duty occurred in 1986 when the company was not only solvent but also enjoying prosperity. The property at risk in reality, therefore, was his and his wife's. The issue is not whether one accepts these mitigating factors as being relevant. Rather, what matters more is that in fashioning partial relief for the director, Hoffmann LJ sought to strike a fair balance by having regard to the interests of the company's outside unsecured creditors while giving due cognisance to the fact that Mr D'Jan suffered significant personal prejudice from his breach of duty. Conversely, if the director is found to have exposed the creditors to too great a risk the court is likely to deny relief. Hence in *Re Simmon Box (Diamonds) Ltd* the court was understandably unwilling to relieve the first respondent who was the active director found liable for negligently losing uninsured gemstones while abroad. Peter Smith QC observed:

> In my view there are no circumstances that can entitle the first respondent to seek relief in this case. First, the conduct was grossly negligent …; second, it was entirely at the risk of the creditors and not just at his own personal risk, and he knew that; and third, he was the prime cause of the loss.[101]

One may see factors, such as the economic realities, as subjective elements of reasonableness. But it would, in our view, be better to recognise them as falling within the range of circumstances that can properly enable the court to exercise the discretion inherent in the notion of fairness. As such in determining whether in fairness relief ought to be granted, the judges bring to bear value judgments that do not, and ought not be allowed to add to the mire surrounding reasonableness. Failing to keep reasonableness and fairness distinct is evident both in the language and context of the preceding passage from Peter Smith QC's judgment.

IV. SECTION 727: THE ECONOMIC REALITIES OF 'THINKING SMALL FIRST'

The reported decisions suggest that relief is commonly sought in the context of a small corporate enterprise where there is a high degree of convergence between who owns and who manages the company. In our view this forms a significant factor encapsulated within the notion of looking to the economic realities when deciding whether or not it is fair to grant relief, particularly when allied to the realities of whose financial interests are in jeopardy. In crude terms such thinking recognises that the defendant director can be principal owner, investor, and perhaps even creditor. In such

[101] N 73 above, 288. See also *Hilton International Ltd v Hilton*, n 99 above.

cases he or she is then *de facto* the company.[102] Corporate doctrine aside, it then becomes justifiable and fair to allow relief, particularly if there are no strong contrary indicators. This explains Buckley J's willingness to afford relief in respect of a technical failure to gain the shareholders' sanction to the £9,000 remuneration paid to the controlling director in *Re Duomatic Ltd*.[103] The judge gave as one reason in support of this conclusion the established practice of making such remuneration payments in a similar way during previous years.

Just as it might be argued that this generosity may do some damage to the spirit if not the letter of corporate personality, it may also be a parochial view that overlooks the wider interests of members or unsecured creditors. In some measure it is an answer to such concerns that the discretion allows the court to award partial rather than total relief. As we have already indicated, such was the case in *Re D'Jan*.[104] This illustrates one facet of the attractive flexibility inherent in section 727. Alternatively, the court may decide to recognise wider concerns in making its award of costs. This occurred in *Re Duomatic Ltd*, where Buckley J, allowing full relief from liability in respect of the director's £9,000 remuneration, found him liable to defray the liquidator's costs. If the interests of either members or unsecured creditors are found to be paramount then the court may decide against awarding relief.[105]

Such pragmatism in utilising the discretionary jurisdiction in section 727 finds resonance with the core policies that underpin the proposals contained in the Steering Group's reports and the consequent White Paper. These are the 'think small first' approach to regulation; the need for an inclusive, open and flexible regime for company governance; and a flexible and responsive institutional structure for company law.[106] The first two in particular acknowledge the need to fashion a corporate law regime that is sensitive to the economic importance of small owner-manager companies. In addition to making the applicable law more accessible and transparent, thinking small may place at a premium the ability to relieve directors from breaches that are technical and only serve to shift financial damage onto the shoulders of small entrepreneurs. The twin goals of accessibility and transparency underlie the Statement of Directors Duties found in Part II of the Government's White Paper II.[107] Clause Four adopts the Law

[102] See the approach taken by Lord Hoffmann in *Meridian Global Funds Management Asia Ltd v Securities Commission* [1995] 2 AC 500, 506–7.
[103] N 38 above.
[104] See n 100 above and associated text.
[105] See the approach of Lord Dunpark, in the Outer House, in refusing relief in order to be fair to the members of the company: *Gibson's Executor v Gibson* 1980 SLT 2; and *Selangor United Rubber Estates Ltd v Cradock (No 3)*, n 51 above.
[106] Final Report, n 4 above, para 1.52; White Paper I, n 4 above, paras 1.2–1.3 and 6.2–6.5.
[107] N 4 above. Schedule 2 sets out the general principles by which directors are bound. Providing a statutory statement found support in the leading empirical survey undertaken for the Law Commissions by S Deakin and A Hughes: see, Law Com No 261 and Scot Law Com 173; n 4 above, 226.

Commissions' recommendation to assimilate into the proposed new companies legislation the objective/subjective test as determinative of the general common law duties of care and skill of directors.[108] For the owner-manager who sees his or her role as director as secondary to his her role as owner, this may predictably raise awareness and consequent anxiety over increased liability in negligence. It is in this context that section 727 has the potential to serve as a mitigating device, the need for which is borne out by empirical evidence prepared for the Law Commission.[109] To meet the concerns of directors of larger companies, whose anxieties over increasing responsibilities may have already been aggravated by *Re D'Jan's* assimilation of section 214 of the Insolvency Act 1986, the Law Commission speculatively suggests that the incidence of D & O insurance is likely to increase.[110] Their prediction may gain even greater force if the company law reforms succeed in raising awareness among directors of what their duties are.[111] Even if insurance take-up does follow this predicted trajectory amongst large companies, in the absence of it being made mandatory, it is far less likely that such policies will be widespread among the small owner manager entities. Directors of such companies may, in any case, find the premiums prohibitive given the risk they pose as compared to their brethren in large companies who are invariably supported by in-house counsel. The continuing absence of D & O insurance, particularly amongst directors of small companies, accentuates a significant lacuna that the maintenance of a revamped section 727 may valuably fill.

V. SECTION 727 AND FIDUCIARY DUTIES

Section 727 will assume greater prominence in the light of the Steering Group's recommendation that section 310 of the Companies Act should be simplified so as to cover only the directors' obligations found in its proposed Statement of Director's Duties. The point is made that the Statement will itself indicate the extent to which any liability arising under its component

[108] *Ibid*, para 5.20.
[109] *Ibid*, 226.
[110] *Ibid*, para 5.10. For a more sceptical view, premised in part on the unlikely rise in the UK of US style class actions against directors, see J Barlow, 'The Risks of the Boardroom' (1997) 5 *International Insurance Law Review* 339, 340. On D & O insurance generally, see V Finch, 'Personal Liability and Corporate Control: The Role of Directors' and Officers' Liability Insurance' (1994) 57 MLR 880 and C Baxter, n 27 above. Indeed, a key recommendation of the recent Higgs report, *Review of the role and effectiveness of non-executive directors* (London, The Stationary Office, 2003) is that companies should indemnify non-executive directors against legal action, and ensure they have appropriate insurance cover (paras 14.14–14.20).
[111] A declared objective underlying the proposed codification of directors' duties is to make the rules accessible to the layman. The White Paper I, n 4 above, para 3.2, cites a 1999 survey of members of the Institute of Directors, which showed that many company directors were ignorant of their general duties.

duties may be varied or waived by the constitution, the board or the general meeting. This has the commendable aim of resolving the apparent contradiction between Table A, article 85 and the terms of section 310. From the convoluted reasoning of Vinelott J in *Movitex Ltd v Bulfield,*[112] it followed that the no-conflict rule was a 'disability' and therefore capable of modification by the constitution. This will no longer hold true. Accordingly, the availability of relief may assume a more prominent status where the breach in question involves directors' fiduciary duties.

There might be an understandable air of anxiety about any enlargement of the remit of the relieving provision into the sacrosanct realms of fiduciary duties. In particular there could be concern about relief encompassing breaches of the cardinal duties of loyalty.[113] After all it is unusual, if not unprecedented, for fiduciary breaches to be either honest and/or reasonable.[114] This, however, begs the prior question of whether or not section 727 is already available for breach of fiduciary duty. There is some Commonwealth support for the view that fiduciary duties are, as a matter of principle, outwith section 727 on the basis of its language, because the provision lists the types of breach that may attract relief;[115] wording which the Steering Group recommend should be axed.[116] The essence of the argument is that where an account of profit is claimed because of a breach of fiduciary duty then there is no loss. Therefore, for the purposes of section 727 there is no liability in respect of negligence, default, breach of duty or breach of trust. That this analysis overlooks the fact that these days the claim may be for equitable compensation rather than for an account of profits has in any event now been overtaken by the decision in *Coleman Taymar Ltd v Oakes.*[117] Judge Robert

[112] [1988] BCLC 104; noted L S Sealy (1987) 46 CLJ 217. To support his view the judge drew upon *Tito v Waddell (No 2)* [1977] Ch 106.

[113] See, most notably, Millett LJ's pronouncement in *Bristol and West Building Society v Mothew* [1998] Ch 1 18: 'Breach of fiduciary obligation, therefore, connotes a disloyalty or infidelity. Mere incompetence is not enough. A servant who loyally does his incompetent best for his master is not unfaithful and is not guilty of a breach of fiduciary duty.' See further, RC Nolan, 'Conflicts of Interests, Unjust Enrichment, and Wrongdoing' in WR Cornish, R Nolan, J O'Sullivan and G Virgo (eds), *Restitution Past, Present and Future: Essays in Honour of Gareth Jones* (Oxford: Hart Publishing, 1998) 87, 88: 'Not all those duties which affect someone properly called a fiduciary are themselves fiduciary duties. For example, duties of care and skill are laid on the trustees and on company directors, yet such duties are not fiduciary duties, albeit that they affect those properly described as fiduciaries. Fiduciary obligations promote loyalty by prohibiting disloyalty, and activity which might lead to disloyalty: fiduciary obligations are proscriptive in nature, and do not encompass the positive duties laid on those described as fiduciaries' (footnotes omitted).

[114] One possible high profile exception is *Regal (Hastings) Ltd v Gulliver* [1967] 2 AC 134n. The relieving provision (in Companies Act 1948, s 448) was not pleaded by the directors in that case: see J Birds, 'The Permissible Scope of Articles Excluding the Duties of Company Directors' (1976) 39 MLR 394, 397; and J Birds, n 12 above, para 10. 2.

[115] *Hytech Builders Pte Ltd v Tan Eng Leong* [1995] 2 SLR 795, 806.

[116] In their Final Report, the Steering Group takes the view that the words 'negligence, default, breach of duty or breach of trust' are an 'indefinite and antiquated formulae'; n 4 above, para 6.3.

[117] N 15 above.

Reid QC rejected as unconvincing this line of statutory interpretation that seeks to put fiduciary breaches beyond the realms of section 727. He went further by allowing relief to the director, Oakes, in respect of the account of any profit arising from a technical breach of fiduciary duty, where he had honestly and reasonably launched another company, GoGas, as a competitor while still technically a director of, but after the termination of his employment with, Taymar. It was fair to award relief because at the relevant time Oakes had been a director in form rather than substance. He had neither been paid nor treated as a director and his former employers had effectively delayed the date of his departure. However, the judge also found that Oakes had acted unreasonably in making unauthorised use of confidential information to negotiate for his own benefit leases that Taymar no longer intended to use. Oakes was therefore held not to be eligible for relief in respect of these breaches of duty as a fiduciary and employee. Judge Robert Reid QC therefore deployed section 727 in a balanced and sensible way, to ameliorate the rigour of fiduciary rules which are often applied in an all too absolutist manner.[118]

On balance, the better view is that there is no absolute prohibition on relief for breaches of fiduciary duties; nor should there be. The point might be put beyond doubt by redrafting section 727 so that it explicitly refers to breaches of fiduciary duty. In any event, where the duty involves questions of loyalty and fidelity, it seems improbable that a director will normally be able to satisfy any or all of the current legislative requirements of honesty, reasonableness, and fairness.[119] However, as *Coleman Taymar* reveals, there are sometimes disputes in which a fiduciary may deserve relief where he or she has merely infringed a fiduciary duty in a technical way. This would also allow section 727 to come to the aid of the entrepreneurial fiduciary without either creating a charter for widespread corporate abuse or risking the integrity of the prophylactic fiduciary rules governing a director's conduct. The value of a relieving jurisdiction in such circumstances may be accentuated by the uncertainty over a director's entitlement to claim an equitable allowance for his or her contribution to making a profit.[120]

[118] Contrast G Jones, 'Unjust Enrichment and the Fiduciary's Duty of Loyalty' (1968) 84 LQR 472 with J Lowry and R Edmunds, 'The Corporate Opportunity Doctrine: The Shifting Boundaries of the Duty and its Remedies' (1998) 61 MLR 515.

[119] This is consistent with Browne-Wilkinson VC's reasoning at first instance in *Guinness plc v Saunders* [1988] BCLC 43, 52. On appeal, the point was left open by Fox LJ in the Court of Appeal ([1988] 1 WLR 863, 871); and found to be irrelevant to the determination of the claim by the House of Lords ([1990] 2 AC 663). For the judicial assumption that relief is available for fiduciary wrongs, see the unreported decisions in *Pleasurebrews Ltd v Vickerstaff* (Court of Appeal), Transcript Association, 17 November 1989, *Re Pantone 485 Ltd; Miller v Bain* (Chancery Division, Companies Court), (Transcript), 29 November 2001; and *Zemco Ltd v Jerome-Pugh* [1993] BCC 275. See further, E. Ferran, *Company Law and Corporate Finance* (Oxford: OUP 1999), 203.

[120] See the House of Lords in *Guinness plc v Saunders, ibid*, 694 (Lord Templeman), 701 (Lord Goff); and *Warman International Ltd v Dwyer* (1995) 182 CLR 544. On this point and the

Overall, it hardly seems likely that this relatively small measure of protection afforded for inadvertent breaches of fiduciary duties will be substantially disturbed even if either or both of the current legislative pre-conditions to relief are abolished. Nor will enshrining the fiduciary duties in the statement of directors' duties necessarily matter. If fairness becomes the lynchpin for the judicial determination of relief, this is more than appropriate and effective as a mechanism to determine whether or not relief should be made available; for it will seldom be fair to assist a dishonest or unreasonable fiduciary.

VI. RELIEF AND CRIMINAL PROCEEDINGS

Almost all the judicial discussion about section 727 and its predecessors in England and Wales has been about its availability in civil proceedings. However, in two decisions, the possibility of relief has been canvassed in respect of a director's criminal liability under section 141 of the Companies Act 1929, where it was alleged that a person had acted as a director without holding the share qualification required by the company's articles. In *Re Barry and Staines Linoleum Ltd*,[121] Maugham J had no hesitation in exercising the relieving provision (contained then in section 372 of the Companies Act 1929) in relation to liability of a 'penal character' in the form of a fine which might not exceed £5 for each day the defendant was in default. He indicated the following robust view of the ambit of relief:[122]

> It is beyond doubt that it applies also where proceedings are being taken in a Court of summary jurisdiction to recover one of the penalties imposed on directors and others under the Act and accordingly, it includes power to relieve against the penalty imposed under s 141.

Although Crossman J declined to relieve directors from the same penal liability in *Re Gilt Edge Safety Glass Ltd,*[123] the learned judge did not question Maugham J's conclusion that relief was available in principle. The proceedings arose because the Bow Street magistrate adjourned the hearing of the summons under section 141 to enable the directors to apply to the High Court for relief. As has already been observed,[124] relief was forthcoming under section 372(2) in respect of any potential future civil liability. However, the view was expressed that, on its correct interpretation, relief from the criminal penalties under section 372(1) could only be exercised by

differences between relief and the equitable allowance see, J Lowry and R Edmunds, n 35 above.

[121] N 71 above.
[122] *Ibid*, 232.
[123] N 70 above.
[124] See text associated with n 70 above.

the magistrates' court in which the proceedings had been commenced. In reaching this view Crossman J apparently found no difficulty, as a matter of statutory interpretation, in accepting that 'proceedings' under section 141 must fall within the expression contained in section 372(2) 'any claim … in respect of any negligence, default, breach of duty or breach of trust'.[125]

Taking such a generous and flexible approach to the statutory language has not commended itself in overseas jurisdictions. There is much to be said for the incisive rigour with which subsequent Commonwealth authorities have roundly criticised and rejected these English judicial pronouncements. In *Lawson v Mitchell,*[126] the authoritative status of both English High Court judgments is challenged because they are delivered *ex tempore,* and because Maugham J's unreasoned conclusion in *Re Barry and Staines Linoleum* is matched by a paucity of argument from counsel. More positively the Supreme Court of Victoria's decision advances detailed arguments to justify confining statutory relief to civil liability. In addition to arguing that the language of the provision is inconsistent with criminal proceedings, the court raises procedural obstacles to its application. Thus, the Court contends that relief is out of kilter with criminal procedures such as jury trial, and that the legislation fails to identify who the respondent is meant to be. Moreover, it is argued that relief is essentially otiose, either because company penal provisions commonly allow 'reasonable steps' defences or attract sliding scale penalties. From this perspective, such discretionary provisions therefore serve the function of relief. Historical analysis forms a final line of attack. It is concluded that the legislative history of the English statute, and its genesis by reference to the power to relieve trustees from liability supported the conclusion that it is intended to apply exclusively in the realm of civil liability. These ostensibly formidable objections to allowing relief in criminal law proceedings have attracted support before the High Court of Singapore in *Re Ideaglobal.Com Ltd.*[127] Addressing the comparable Singaporean provision in section 391(2) of the Companies Act (Cap 50), Lee Seiu Kin JC rejected the English decisions because:

> The decision of the Full Court of the Supreme Court of Victoria in *Lawson v Mitchell* is the only decision in which a full consideration was made of the scope of the provision. In my respectful view, the reasoning there is unassailable.[128]

This may exaggerate the calibre of the reasoning. For one thing, it is not easy to discern whether or not the architects of the 1907 Act intended relief to be available for a director's penal liability. The 1907 Act itself introduced

[125] *Ibid*, 501. Although it was necessary for the Court of Appeal to pronounce on this question in *Customs and Excise Commissioners v Hedon Alpha*, n 83 above, Ackner LJ's judgment, at 823, offers passing and indirect endorsement for the view in the two earlier English cases.
[126] N 26 above.
[127] (2000) 3 SLR 100.
[128] *Ibid*.

new penal provisions.[129] There is also one contribution to the parliamentary debates from which it might be concluded that there was some contemporary expectation that relief should be available in such circumstances.[130] Where *Lawson v Mitchell* is unassailable is in its forensic assessment that the language in section 727 does not best accommodate criminal proceedings. Of course this might be rectified through careful re-drafting. What is more pressing is the policy question: should directors be able to seek relief where their conduct falls foul of the criminal law? Or, is this a step too far? *Lawson v Mitchell* may be right to point out that the necessity for such intervention is ameliorated where discretion exists as to penalty. Equally the *mens rea* component of many offences may make it difficult for defendants to show that their offence is honest, reasonable and fair. As *Re Gilt Edge Safety Glass Ltd* indicates, sometimes a director's breach may be purely technical and in such circumstances there does not seem to be any harm in allowing relief even if this is an additional mechanism by which he or she may be protected partially or fully from the penal consequences of a minor transgression. This will not pass the cost onto the company nor will it damage the prophylactic purpose of such penalties. Against this liberal approach to the ambit of relief, it has to be noted that the status and role of criminal sanctions figures prominently on the current company law reform agenda which proceeds on the basis that offences of dishonesty should be strengthened.[131] If this leads to an accentuation in the way the law should differentiate between the commission of company offences that are technical as opposed to dishonest, then it may effectively render it less necessary to include relief for criminal matters in the successor to section 727.

VII. CONCLUSION

Understandably the preponderance of textbook wisdom casts section 727 as something of a Cinderella provision in company law. A cursory examination of its history and the paucity of successful applications over the course of its statutory lifetime does little to disturb this impression. Although the readily available historical accounts of what inspired the introduction of the progenitor of section 727 may be scant, the architects drew unquestionably upon the blueprint provided by the contemporary introduction of a discretionary relieving power in favour of trustees. This transplantation is itself hardly surprising in view of the fact

[129] Likewise, as Stephenson LJ observes in fathoming the meaning of the word 'default' in s 372 of the Companies Act 1929, that statute introduced twenty or so new penalties affecting company officers: see *Customs & Excise Commissioners v Hedon Alpha Ltd,* n 83 above, 825.
[130] See Mr Lupton MP (Lincolnshire, Sleaford), n 34 above, 897.
[131] See Rt Hon Lady Justice Arden, DBE, n 4 above, 597.

that both the company and trust law review bodies shared the same chairman in Sir Robert Reid. At first sight the rationale for the judicial relieving provision is nowhere spelt out in that committee's report. However, its focus seems to have been upon injecting fairness by means of a discretionary relieving provision which might serve as a mitigating device, offsetting liability for breach of statutory duties arising, for example, in relation to company prospectuses: duties which the committee did not think either 'safe or wise to diminish'.[132] It will be recalled that one member of the committee, Mr Speyer, harboured broader concerns that can still be found echoing in the most recent review of the future of company law reform commissioned by the DTI and culminating in the White Paper. These are a desire to foster economic enterprise and a need to address disquiet over the lack of a statutory formulation of an appropriately demanding standard of care for directors. Similar anxieties colour the current reform agenda. Moreover, it is as much these concerns, as opposed to the severity of statutory liabilities, that can be discerned as being amongst the factors that have figured prominently in the judicial interpretation of the language and scope of section 727, even if their influence has not always dictated a uniform approach to the key requirements of honesty, reasonableness and fairness.

The origins of the provision, and the judicial views expressed about it, therefore hold significant insights when considering the continuing value and role of a statutory relieving provision. They can also offer an antidote to any argument that the low number of successful claims recorded in the law reports best measures the worth of the provision. It may well be that the section is commonly invoked by defendant directors alongside other defences as a 'belt and braces' tactic.[133] This may go some way to explain why its citation is often peripheral, seldom leading to extensive judicial consideration of the language and ambit of the provision. However, the volume of successful litigation should not of itself become the definitive yardstick by which its past, present or future value should be measured.[134] Its inherent value has been accepted by Parliament, which has extended the reach of section 727 beyond the sphere of the limited company.[135] Along with its trust's law counterpart, the provision has provided a blueprint for the

[132] N 16 above, para 24.

[133] *Joint Receivers and Managers of Niltan Carson Ltd v Hawthorne* [1988] BCLC 298, 328 (Hodgson J).

[134] It has been noted that, although the jurisdiction is broad, 'relief is not lightly given': Law Com (261), n 4 above, para 11.42. See also E Ferran, n 119 above, 204 n 337.

[135] Thus for the purposes of s 727 the British Wool Marketing Board is to be treated as a company and its members as officers (Agriculture Act 1993, s 57); and eligibility for relief has been extended to include an auditor or independent examiner appointed by a charity (Charities Act 1993, s 44) and an officer or auditor of a friendly society (Friendly Societies Act 1992, s 106)

proposal for an equivalent statutory safeguard for people serving on the governing bodies of public service organisations.[136]

Even if the precise fate of section 727 has always been marginal, equally there is no pressure for its abolition. Whatever company law changes are made in the twenty-first century, the case for some mechanism offering directorial protection from personal liability for breaches is unlikely to diminish. Indeed, one justification for relief for 'technical' breaches of duty may lie in the fact that this jurisdiction has not explicitly embraced a North American style business judgment rule. Without suggesting that relief and the business judgment rule work identically,[137] there are occasions when the award of relief for liability can be seen as rooted in considerations which, under the safe-harbour rule, might have prevented liability being engaged in the first place.[138]

A significant challenge may be to craft a provision that maintains a sufficient check against relief for liability becoming a charter for abuse of the corporate entity. The danger is not simply that directors who are too readily insulated from liability may go on to repeat their mistakes; their original mistake may even have an adverse impact upon the national economic well-being. Yet there can be few who would wish to see the power to award relief disappear in cases where the director's error is technical, particularly if he or she, as a sole or substantial owner of the enterprise, is the only person whose financial interests are prejudiced. It may assist in delivering the 'think small first' aspirations adumbrated in the White Paper. It has been our contention that this admirably illustrates the manner in which fairness is the lynchpin of section 727. This is not to advocate that section 727 should have no application to a director of a public company with diverse ownership.

The real question is not whether such a provision should continue to exist but in what form. Doubtless there are a number of possible ways in which the provision might be re-worked. Although the case law has thrown up more questions about the ambit of the legislation than definite answers, the judicial treatment of section 727 (and its predecessors) has identified a number of critical issues facing the draftsman about the gateway preconditions of honesty and reasonableness and the concept of fairness that is the essence of the discretionary power. As this chapter has indicated, these concern the degree to which the tests of honesty and reasonableness should be imbued with objective/subjective characteristics; the inter-relationship

[136] See E Hambley *et al, Personal Liability in Public Service Organisations: A Legal Research Study* (London: The Stationary Office, 1998) a report prepared for Lord Neill's Committee on Standards in Public Life.

[137] See MR Pasbarn *et al*, n 3 above; and J H Farrar, 'Towards a Statutory Business Judgment Rule in Australia' [1998] *Australian Journal of Corporate Law* 302.

[138] This accords with one commentator's interpretation of *Re Claridge's Patent Asphalte Co Ltd*, n 98 above: See HJ Ford, RPA Austin and IM Ramsey, *Ford's Principles of Corporations Law* 10th edn, (Australia: Butterworths, 2001), para 8.420.

between section 727 and other statutory regimes (most notably wrongful trading); and the applicability of relief both for breaches of fiduciary duties and criminal offences. It is not easy to see how or why the draftsman should attempt to find explicit solutions to all of these tensions that have surfaced in the case law. However, in respect of the first matter, *Re D'Jan*'s preference for a pragmatic and subjectively biased interpretation of reasonableness is now being judicially questioned.[139] Likewise the courts are gradually recognising that they may need to pronounce upon the extent to which honesty is or is not determined subjectively and/or objectively. To date this has not resulted in a full-blown consideration of the issues that have been canvassed more explicitly in the sphere of breach of trust by the House of Lords in *Twinsectra*. Even if it is desirable or possible to distil the concluded view into legislation, it is our view that this is rendered unnecessary. A better solution might lie in deleting the words honest and reasonable from the statute. This would not make such considerations redundant. Rather these may find expression in the concept of fairness. This will obviate the semantic wrangles on the precise test to be applied to establish either of those pre-conditions. It would arguably make the relieving provision more transparent than at present because under the current law the courts have not succeeded in keeping as neat a division between these three notions as has sometimes been suggested. There is sufficient evidence in the case law to justify the conclusion that fairness is able to balance the protection of the director with the needs of other constituencies, such as members and creditors. Relying upon concepts of equity and fairness is not without statutory precedent,[140] and is the formulae adopted by the White Paper in framing a specific exemption to certain directors' liabilities within the revised capital maintenance regime.[141] In our view it best encapsulates the essence of the past discussion of how the provision is meant to work and best points the way forward.

[139] By the Court of Appeal in *Bairstow,* n 43 above, and text associated with n 75 above. Without reviewing s 727, Law Com No 261, n 4 above, notes: 'We have proceeded on the basis that the interpretation given to the section in *Re D'Jan* will be upheld by later decisions. We consider that it will.'

[140] See, for example, ss 113(7) and (8) of the Companies Act 1985.

[141] See White Paper II, n 4 above draft clause 44(3).

20

Enron and the Long Shadow of Stat. 13 Eliz.

DOUGLAS G BAIRD*

ONLY TWO YEARS ago, Enron was one of the most admired corporations in America. Today its assets have been sold and its managers led off in handcuffs. Creditors have discovered that Enron's assets will give them only a few cents on the dollar. Those closest to fraud and other mischief are likely judgment proof. But a number of entities just off stage have deep pockets and the creditors are in active pursuit of them. In this paper, I set out the challenges that await Enron's creditors and assess their prospects. The creditors have some reason for optimism. Often the best paths are the oldest and the most well trodden, and this case is no exception. While the transactions in Enron involved full-return swaps, derivatives, and every variety of financial instrument, the best course for the creditors is set out in the Statute of 13 Elizabeth, as interpreted by Lord Coke in *Twyne's Case* in 1601.

I. ENRON'S BUSINESS PLAN

Financial analysts became smitten with Enron's business plan.[1] But we should not be too quick to ridicule them. Almost any innovative idea looks bad if it fails. In 1907, cars were a plaything for the rich and paved roads

* An earlier version of this chapter was presented at the Commercial Law and Commercial Practice Seminar at the London School of Economics in November 2002. Its participants and especially Kevin Davis provided helpful comments. Many of the ideas in this chapter grow out of my long, fruitful, and continuing collaboration with my colleague Robert Rasmussen, to whom I am, as always, most indebted. For its support of this and many other projects, I am grateful to the John M Olin Foundation.

[1] For a pre-lapsarian account of the analysts and Enron, see Brian O'Reilly, 'The Power Merchant' *Fortune*, 17 April 2000, at 148 ('In January, before a room packed with Wall Street securities analysts, Enron 'broke radio silence' on its plans .... It was like Jesus showing up at a tent revival. Analysts swooned; they cheered; one declared that Enron had 'instant credibility' in the new endeavor').

were the exception. Nevertheless, a car designer who had already failed twice bet millions on a radically new car. By using exotic vanadium steel alloys and stamped metal casings, he believed his car could be produced for $500 and that millions of average Americans would buy it at that price, even though $500 at the time was the better part of a year's wages. In the 1950, the Haloid Company risked all on a machine that made copies of documents on plain paper, even after the experts at IBM and General Electric had established that there was no market for such a machine as carbon paper was a cheaper and simpler way of making duplicates. In 1986, an entrepreneur in Seattle bet every penny that ordinary Americans would spend several dollars a day on espresso and cappuccino. All these ideas seemed far-fetched in their time. But Henry Ford's Model T changed the world. Haloid thrived (and changed its name to Xerox). Starbucks has become a fixture on every street corner.

The analysts who looked at Enron found a business plan that seemed more plausible, and the potential payoff was enormous. Moreover, those who had bet against Enron over the previous decade had been wrong. Enron was once a sleepy, debt-ridden pipeline that provided point-to-point transport of natural gas. It worked in a heavily regulated environment that placed few demands on its managers. Enron's world changed in the late 1980s when the regulations were lifted. Its managers suddenly realized that Enron's pipeline did not have to work as a discrete link between one gas well and a given end user. Instead, deregulation allowed Enron to use its pipeline as a network. It could reallocate natural gas from where there was excess capacity to where there was excess demand.

Enron also discovered that, once free of regulation, it was able to refashion contracts to purchase and sell natural gas according to the needs of its buyers and sellers. Its network protected it from liquidity shocks that can unseat market makers. In short order, Enron created a 'gas bank' that allowed buyers and sellers to lock in the prices for natural gas.[2] Enron went quickly from being a company that transported natural gas to a nimble market maker that allowed everyone from Owens Corning to the Archdiocese of Chicago to control their energy costs.

As the market for electricity became deregulated, Enron again recognized that, once deregulated, the power grid was another network that allowed market makers to match buyers and sellers. Soon Enron was making a market in electricity. It was an intermediary that could use its network to exploit the gains from trade. A utility in Boston that had more natural gas than it needed and a utility in California that wanted additional electricity cannot trade with each other directly. Enron, however, could create a series of transactions, using its network and its ability as a market maker to effect such a trade.

[2] See Peter C Fusaro & Ross M Miller, *What Went Wrong at Enron* (John Wiley & Sons 2002) 29–34.

After proving itself as a successful maker of markets in gas and electricity, the analysts who once doubted that markets could be made in gas or electricity now believed Enron as it tried to establish markets in other commodities. Enron had something to offer any business subject to risks that it had no skill in managing. Before Enron filled the niche, a newspaper's profits would turn on the changing price of newsprint. The publisher had no way to lock in a reliable supply of newsprint for an extended period. Enron's managers created markets for newsprint, water, weather, and many other commodities.[3] Enron, by its account, had a comparative advantage in creating markets acquired from its years in the energy business.[4]

Enron prided itself on its ability to spot opportunities. During the summer of 2000, Enron helped a zinc producer in the Northwest shut down its operations for 6 weeks and sell the power it would otherwise have used to a buyer who needed it more. Enron helped the zinc producer find the buyer and provided a financial derivative to ensure the sale at a fixed price. Enron also supplied zinc from its metals subsidiary so that the producer could meet pre-existing obligations.[5] Such transactions make everyone better off.

Enron's managers believed its ability to match sellers and buyers of energy would allow it to create exotic new markets. Internet traffic was growing geometrically. Some believed that it was doubling every 100 days.[6] But the Internet itself was not well-managed. At times, people would overload parts of the system and at other times it would remain largely unused. Enron could create a market in broadband, just as it created one in gas and electricity. Its pipelines gave it the ability to lay fibre-optic cable cheaply, and it formed alliances with other firms that would allow easy shifting of the cable from one user to another. Just as it acquired natural gas to put through its pipelines, it could also acquire content to pump through the fibre-optic cable. It reached a deal with Blockbuster to provide video on demand.[7]

The ultimate vision Enron put to the rest of the world was one in which a single firm (Enron) would be where everyone turned to ensure reliable sources of supply and avoid risks that they had no particular ability to manage.

[3] See O'Reilly, above n 1, Fortune, 17 April 2000, at 148.
[4] See Enron Annual Report 1999, at 2 (2000) ('We are clearly a knowledge-based company, and the skills and resources we used to transform the energy business are proving to be equally valuable in other businesses'.); *ibid*, at 5 ('The fundamental skills and expertise we use to develop energy and communications solutions can be applied to many situations that inhibit our customers' profits and growth'.); Enron Annual Report 2000, at 5 (2001) ('We have a proven business concept that is eminently scalable in our existing businesses and adaptable enough to extend to new markets').
[5] See Enron Annual Report 2000, at 12 (2001).
[6] See Yochi J Dreazen, 'Wildly Optimistic Data Drove Telecoms to Build Fiber Glut', *Wall St J*, 26 September 2002, at A1.
[7] See Robert Preston & Mike Koller, 'Enron Broadens into Broadband', *InformationWeek*, 6 November 2000.

Carmakers' profits would turn on the cars they made, not on what happened to the price of palladium.[8] The manager of the office building would know in advance how much it would cost to heat the building during the winter and cool it during the summer.

Enron could exploit economies of scale and enjoy the liquidity that comes from having its own supplies of the gas, electricity, paper, precious metals, and much else. Others could try to imitate Enron, but Enron had a first-mover advantage. The expertise it had gained in making markets ensured it would be better than anyone else. In a winner-take-all market, the firm that is even a little bit better flourishes and the others disappear. Standard Oil became dominant because it could refine and sell oil at a profit even when it sold the oil for less than its competitors' costs. Jeffery Skilling boasted that Enron was in the same position as Standard Oil in 1890.[9]

In short, Enron was a new economy company with a long track record. While the dot.coms never had any earnings, Enron had a solid record of earnings. It had outperformed the market for the better part of a decade. Anyone looking at Enron, of course, could see risks. It had taken a huge one-time write-off in 1997 because of a bad deal involving North Sea natural gas. Its managers were arrogant, and its financials were opaque. But if Enron succeeded as it had in the past, it would play a central role in the new economy.[10]

A large part of Enron's failure can be traced not to any misdeeds, but simply to the failure of its business plan. The analysts who believed in a bright future for Enron made a systematic mistake about the opportunities in the information industry.[11] Analysts misjudged the growth in Internet traffic. While the growth was large, it was far short of the 100 per cent every 100 days. Moreover, new technologies dramatically expanded the capacity of existing network. The effects of the deregulation of the telecommunications industry in 1996 were also poorly understood. Billions invested in hard assets proved worthless. Much of the fibre-optic cable

[8] Ford Motor Company took a $1 billion charge to its earnings to reflect the loss in value of the palladium it had stockpiled for its catalytic converters. See Norihiko Shirouzo, Gregory L White & Joseph B White, 'Driving Lessons: Beyond Explorer Woes, Ford Misses Key Turns in Buyers, Technology', *Wall St J,* 14 January 2002, at A1.

[9] See Christopher Palmeri, 'At the Heart of a Revolution', *Forbes*, 12 January 1998, at 48.

[10] See Erin Davies, 'Enron: The Power's Back On: Rousing a $20 Billion Giant', *Fortune*, 13 April 1998, at 24. Not only did Enron expressly style itself as a 'new economy' firm, see 1999 Annual Report, at 2 ('When you define a New Economy company, you define Enron'.), but its annual reports draw heavily on the new economy lexicon. See, for example, Enron Annual Report 1998, at 3 (1999) ('business platform'), Enron Annual Report 1999, at 2 ('knowledge-based company', 'global networks', 'What you own is not as important as what you know,' 'constant innovation,' 'connectivity,' 'strategic contractual relationships'); *ibid,* at 4 ('first mover advantage', 'leverage'); *ibid,* at 5 ('intellectual capital'); Enron Annual Report 2000, at 2 ('[r]obust network of strategic assets'), *ibid,* at 3 ('integrating EnronOnline into all our businesses as an accelerator'), *ibid,* at 4 ('network connectivity'); *ibid,* at 5 ('leverage').

[11] See Dreazen, above n 6.

laid was ‘dark’. It has never been used and never will be. Trillions in market valuations have disappeared. The valuation of Enron was premised in part upon its ability to take advantage of the shortage in broadband. When the shortage disappeared, a large component of Enron’s value in the market disappeared as well.

Second, Enron’s business plan required it to be too many things at the same time.[12] It had to be a supplier of basic commodities, it had to be a market maker in those commodities, and it had to make the alliances and strategic investments all at the same time. It was as if a single firm was an oil refiner, a stock exchange, and a venture capitalist. Enron could not maintain its trading operations where it routinely entered into 10-year contracts unless it appeared to be financially sound. It needed to meet its earnings targets and maintain its credit-rating. Its crown jewel, its trading operation, would collapse without it. But Enron’s business plan required it to make bold investments that might not bring returns for many years, if at all. Put differently, Enron was a firm that could maintain its high stock price only by making bold and risky moves, but it could maintain its on-going operations only if its earnings were reliable and steady. These two forces are fundamentally at odds.

Beginning in 1997, Enron began taking steps to ensure that outsiders believed that its operations were sound and that it was a reliable counterparty. Instead of insulating its trading activities and making sure that its fortunes were independent of the rest of the firm, Enron resorted to a series of increasingly elaborate transactions that gave the appearance of solid earnings and ensured that its bond rating would remain high enough to maintain itself as a reliable counterparty. These arrangements made it appear as if the dot.com collapse left it unaffected. As a result, its stock price continued to rise as other firms closely tied to the new economy fell.

The expertise Enron developed in managing its network and creating markets served it well in creating subsidiaries and Special Purpose Entities (‘SPEs’) that fooled investors, analysts, and rating agencies. Put in its simplest form, imagine (in a world in which the discount rate is 0 per cent) that Enron wants to borrow $100 from Investor, but it wants to keep the loan hidden. Enron can sell $100 of stock in Investor. At the same time, one of Enron’s 4,000 subsidiaries can enter into an arrangement with a subsidiary of Investor in which the Enron subsidiary acquires the right to buy the stock back for $100 at a specified time. Another subsidiary gives Investor the right to sell the share of stock to it for $100. If one were able to view all these transactions together, we would see that Investor had the same economic position vis-á-vis Enron as a creditor. It gave Enron $100 and returned in exchange a right to get back $100 at some time in the future.

[12] See Douglas G Baird & Robert K Rasmussen, ‘Four (or Five) Easy Lessons from Enron’, (2003) 56 *Vand Law Review.*

The amount it would receive from Enron would be the same whether Enron fared well or poorly. It enjoyed neither the upside nor the downside we associate with equity. But outsiders will know this only if they can step these transactions together. In a world in which Enron itself is engaged in massive trading activities as its day-to-day business, outsiders have no way of doing this.

The creditors need to identify those transactions involving SPEs in which assets left Enron, still have value, and can be retrieved from the third parties that are now holding them. Nearly all these transactions involve the use of derivatives and other financial instruments. Nothing about the legal theory available to Enron's creditors turns on this, however. Hence, the easiest way to understand them is to step away from such exotic manoeuvres and focus instead on mundane transactions in which a firm engages in transactions that have the same form.

II. RAPTORS AND OTHER BEASTS

One of the representative transactions that Enron used is Raptor III. Enron created a subsidiary, The New Power Company (TNPC), in which it owned a 75 per cent interest. TNPC was designed to provide energy to retail customers in deregulated markets. Enron then engaged in a set of transactions designed to reflect currently future revenues from this business. Enron transferred TNPC stock to a special purpose entity (Hawaii 125–0) and booked large gains as it marked the value of the stock to market. But along with this transfer, Enron also engaged in transactions involving total return swaps that exposed it to risks of subsequent changes in the value of the TNPC stock, and mark-to-market accounting would also require it to incorporate any subsequent decline in the value of TNPC stock in its earnings. The purpose of the Raptor III was to be able to show income that would offset these losses.

Enron used Raptor III to avoid showing a loss if TNPC fell in value. To do this, Enron used a limited partnership (LJM2) run by its Chief Financial Officer (Andrew Fastow) to 'hedge' (at least from the perspective of its income statement) the losses it might incur from declines in the value of TNPC stock. LJM2 became the principal shareholder of a limited liability company (Porcupine). Porcupine acquired TNPC stock (through yet another entity) from Enron and gave it a promissory note in return. Porcupine then engaged in a hedging transaction with Enron that obliged it to pay Enron in the event that TNPC stock went down. Within a week, Porcupine also made a distribution to LJM2 that returned to it the entire amount it had invested in Porcupine. Porcupine had no assets other than TNPC stock.

Porcupine would be unable to repay its note to Enron if the stock went down. Moreover, if the stock went down, it would have no ability to honour its obligations under the hedge. The transaction allows Enron to prevent declines in the value of TNPC from affecting its earnings over the short term, but over the long term, the transaction brings Enron no benefit. Indeed, the transaction is affirmatively costly. Not only are there fees to lawyers and accountants, but the investors in LJM2 earned extraordinary returns. In chapter 11, the creditors of Enron now need to figure out Enron's claims against Porcupine, its rights to whatever assets Porcupine still has, and its rights against LJM2 and any others who invested in Porcupine or transacted with it. The bankruptcy questions raised by Raptor III replicates itself in many different guises in Enron.

The transactions used elaborate hedges and derivatives and were motivated by arcane accounting rules, but the legal problem can be separated from them. We can focus on them by using the following hypothetical. OfficeCo builds and manages office buildings. It has shown impressive growth and consistent earnings in a highly volatile real estate market. Much of OfficeCo's success is attributed to, in the financial press's words, its 'ingenious and creative' CFO. One of OfficeCo's recent projects was carried out by its wholly owned subsidiary, White Elephant Enterprises. Its sole asset is White Elephant Plaza, a large office building that cost \$800 to build, much more than anyone expected. OfficeCo manages the building. The managers of OfficeCo believe that the Plaza is worth at least \$1000, but they fear it may decline in value. To protect OfficeCo's track record, they decide to sell the Plaza. To their surprise, they cannot find anyone willing to buy the property for \$1,000. Indeed, they cannot find anyone willing to buy it from them for \$800.

The CFO, however, finds a way to ensure that the uncertain fortunes of the Plaza do not jeopardise OfficeCo's solid reputation among investors. With the approval of the Board, the CFO creates CFOPartners, a limited partnership. The general partner of CFOPartners is the CFO and the limited partners are a group of outside investors. The CFO contributes \$2 and the limited partners put in \$198.

CFOPartners then forms WhiteElephantCo, a wholly owned subsidiary of CFOPartners. WhiteElephantCo buys OfficeCo's equity stake in White Elephant Enterprises for \$1000 (\$200 in cash and a long-term unsecured note for \$800). WhiteElephantCo also enters into a long-term contract with OfficeCo that gives OfficeCo the right to manage the building. Soon after the purchase of the Plaza, WhiteElephantCo borrows \$250 from Bank and gives it a first mortgage on the Plaza. A little later, WhiteElephantCo declares a dividend of \$250. CFOPartners distributes the cash to the partners.

OfficeCo reports record profits for the year, and its CFO continues to be highly praised for his astute and aggressive management. Three years pass, but then OfficeCo's fortunes take a turn for the worse. OfficeCo defaults on its loans. The CEO and CFO are fired. The firm files for bankruptcy. The assets of OfficeCo are only enough to give its general creditors a return of 10 cents on the dollar. The creditors of OfficeCo would like to bring White Elephant Plaza into the bankruptcy estate free of Bank's security interest. They would also like to recapture the $250 paid out to the investors in CFOPartners.

The creditors of OfficeCo are fundamentally in the same position as the creditors of Enron. Enron took an illiquid asset (TNPC stock) and transferred it to a Special Purpose Entity (Porcupine) in a way that protected it from the risk that its value would fall. OfficeCo took an illiquid asset (White Elephant Plaza) and transferred it to a Special Purpose Entity in a way that also allowed it to lock in the value of the asset. In both cases, the outside investors that made the transaction possible (LJM2 and CFO Partners) enjoyed substantial returns while putting nothing at risk. The general legal theories that are relevant to solving these problems are the same in both cases. We explore these in the next part of this paper.

III. AVENUES OF RECOURSE

When we take several steps back from this hypothetical, we can see that something is amiss in what OfficeCo has done. When we look at the various discrete transactions as a single deal, we can see that OfficeCo is spending real resources merely to give outsiders the illusion of financial stability.

Because WhiteElephantCo has no assets of its own, the transaction does nothing to protect OfficeCo if Plaza declines in value (or simply turns out to be worth less than the amount of the note, an amount that exceeded what any third party was willing to pay for the property). By selling the asset to WhiteElephantCo, any upside will be enjoyed by CFOPartners. OfficeCo has given up its right to enjoy the upside in the event that the real estate market improved.[13] The principal beneficiary of this transaction is CFOPartners. When the transactions are stepped together, we see that it put nothing at risk, and ended up with $50 in cash and all the upside in the event that Plaza is ever worth more than $1000. Though not couched in these terms, CFOPartners ends up, in addition to the cash, with the equivalent of an option to buy Plaza for $1000.

Anglo-American debtor-creditor law relies in large measure on creditors to protect themselves. Many loan agreements would prohibit the sale of

[13] In some Enron transactions, an entity related to Enron engaged in a total return swap with the special purpose entity that would ensure that Enron would enjoy the upside in the event that the asset rose in value. In the context of the hypothetical, such an arrangement would have the effect of OfficeCo acquiring an option to buy White Elephant Plaza for $1000. OfficeCo's acquisition of this option is as troubling as giving the upside to CFOPartners. If OfficeCo retains it, then the transaction is utterly without economic substance. If OfficeCo gives it up, it is parting with part of the economic value of the asset and receiving nothing in return. As discussed below, each is a different, but equally compelling 'badge of fraud' that makes the transaction suspect.

Plaza without the creditor's blessing. Our concern, however, is with the background protections that creditors as a group enjoy in bankruptcy. Such protections are necessary because not all creditors enjoy written loan agreements. Moreover, even the most elaborate loan agreements cannot enumerate all the ways in which a debtor set upon mischief can compromise their rights. We need basic principles that work in simple cases (such as ours) and in more elaborate ones (such as those we see in Enron).

A. Caplin's Legacy and Piercing the Corporate Veil

Individual creditors may be able to bring actions based upon non-bankruptcy law, such as a violation of the securities laws, against CFOPartners, Bank, and the investors. Such causes of action are not sure-fire, even if the transactions, when viewed as a whole, seem suspect. For example, the Supreme Court has held that private plaintiffs cannot bring Rule 10b–5 actions against those who merely aided and abetted violations of the securities law.[14] Showing that the various third parties were themselves directly responsible is harder. State law theories of liability exist as well. In addition to state blue sky laws, various common law theories of recovery are available. Transactions less conventional than OfficeCo and WhiteElephantCo might be ultra vires or illegal.[15]

Our focus, however, is not on the discrete actions that might be available, but rather on those that can be asserted on behalf of the creditors in bankruptcy. It might seem that CFOPartners and WhiteElephantCo are merely empty shells that are impermissible manipulations of the corporate form. Non-bankruptcy law sometimes authorises creditors to disregard the corporate form and 'pierce the corporate veil' in such situations. These entities are not truly separate, but rather merely 'alter egos' of OfficeCo. All of the power over White Elephant Plaza resides in the CFO of OfficeCo. OfficeCo makes all decisions about the Plaza and continues to manage it as it did before the sale. WhiteElephantCo is nothing more than a set of bookkeeping entries. To go beyond mere form, one should ignore WhiteElephantCo and treat the Plaza just like all of OfficeCo's other assets. This theory, however, immediately runs up against a well-established doctrine of US bankruptcy law.

Section 544(b) of the Bankruptcy Code was written in the shadow of *Caplin v Marine Midland Grace Trust Company*.[16] The trustee can avoid only *transfers* that the debtor made or obligations the debtor incurred. It does not give the trustee the right to bring *damage actions*. The trustee can

[14] *Central Bank of Denver v First Interstate Bank of Denver*, 511 US 164 (1994).
[15] See *In re Adler Coleman Clearing Corp*, 263 Bankr 406 (SDNY 2001).
[16] 406 US 416 (1972).

bring damage actions only if the debtor itself would have been able to bring the action outside of bankruptcy and that are therefore property of the estate within the meaning of §541.[17] Hence, the estate can bring a veil-piercing action on behalf of the creditors only if the debtor would be able to bring the cause of action under non-bankruptcy law.

In our example, the *Caplin* principle limits the ability of the trustee to bring this transaction in the bankruptcy case. The estate can bring a veil-piercing action, like any other damage action, only if the debtor could have maintained the action outside of bankruptcy. Some federal courts have found that the debtor can bring veil-piercing actions against its own shareholders, and hence the trustee can as well because of § 541.[18] State law authority for this proposition, however, is not easy to find. Indeed, the gravamen of the action—that the firm and the shareholders are one and the same—is inconsistent with the idea that the firm can sue the shareholders. At least one state supreme court has repudiated a federal circuit court's opinion that the debtor did possess such a cause of action against its shareholders.[19]

Even if we overcome this obstacle, veil piercing in our example (or in the case of Enron) does not fit comfortably within traditional notions of veil piercing. The shareholder of WhiteElephantCo is not in fact related to OfficeCo. When the alter ego doctrine is invoked, it is usually between a subsidiary and a parent that holds all of its equity. Here WhiteElephantCo is owned by CFOPartners. OfficeCo has no ownership interest in it. The former CFO may run CFOPartners, but it is a distinct entity and it has substantive economic rights (the upside in the event that the Plaza goes up in value).

Even if we could ignore the lack of a formal relationship between the two firms, other problems remain. The typical state veil-piercing action takes three forms: (1) Individuals wholly neglect the formalities of corporate form in running their affairs and those of their sole proprietorship; (2) An entrepreneur sets up a parent corporation with many subsidiaries whose operations are hopelessly intertwined; or (3) Tort victims bring suit against the parent of subsidiaries that have been deliberately undercapitalised. Veil-piercing actions in which creditors of a *parent* (OfficeCo in our case) reach assets in the hands of a subsidiary, however, tend to be less successful.[20]

[17] See, eg, *Mediators, Inc v Manney*, 105 F 3d 822 (2d Cir 1997); *Steinberg v Buczynski*, 40 F 3d 890 (7th Cir 1994); *Schertz-Cibolo-Universal City Independent School District v Wright*, 25 F 3d 1281 (5th Cir 1994); *Williams v California First Bank*, 859 F 2d 664 (9th Cir 1988).
[18] See, eg, *St Paul Fire & Marine Insurance Co v Pepsico, Inc*, 884 F 2d 688 (2d Cir 1989); *Steyr-Daimler-Puch of America Corp v Pappas*, 852 F 2d 132 (4th Cir 1988); *SI Acquisition, Inc v Eastern Delivery Service*, 817 F 2d 1142 (5th Cir 1987). See also *Koch Refining v Farmers Union Central Exchange*, 831 F 2d 1339, 1347 n 11 (7th Cir 1987).
[19] See *In re Rehabilitation of Centaur Insurance Co*, 632 NE 2d 1015 (Ill 1994).
[20] See Robert B Thompson, 'Piercing the Corporate Veil' (1991) 76 *Cornell Law Review* 1036, 1055.

In assessing the strength of a veil-piercing action, the law of the specific state matters.[21] Under Oregon law, for example, the test seems to focus on 'improper conduct' by the party seeking to take advantage of the legal separateness of the entities.[22] Under New York law, the corporate veil can be pierced in the absence of fraud when the 'corporation has been so dominated by an individual or corporate parent that the subsidiary is relegated to the status of a mere shell, instrumentality, or alter ego'.[23] Under Texas law, with respect to contract claims, actual fraud must be shown. Constructive fraud is insufficient.[24]

If veil-piercing actions are unavailable, the trustee might still seem to be able to ask the bankruptcy court to invoke its equitable power to achieve much the same end. The doctrine of substantive consolidation does in bankruptcy what veil piercing does outside.[25] When two bankruptcy estates are substantively consolidated, we combine the assets and liabilities of both.[26] The affairs of the two firms may be so closely entwined that each lacks a separate existence. Substantive consolidation avoids the hard conflict of laws and standing questions that state veil-piercing brings:

> The consolidated assets create a single fund from which all claims against the consolidated debtors are satisfied; duplicate and inter-company claims are extinguished; and the creditors of the consolidated entities are combined for purposes of voting on reorganization plans.[27]

[21] Traditional conflicts rules suggest that the forum jurisdiction will look to the conflict rules of the state of incorporation. See *Fletcher v Atex, Inc*, 68 F 3d 1451, 1458 (2d Cir 1995); *Wausau Business Insurance Co v Turner Construction Co*, 141 F Supp 2d 412 (SDNY 2001). But we do see cases in which courts apply the law of the forum. See, eg, *Vuylsteke v Broan*, 17 P 3d 1072, 1074 n 2 (Oregon App 2001).

[22] See *Amfac Foods, Inc v International Systems & Controls Corp*, 654 P 2d 1092 (Oregon 1982). Enron is incorporated in Oregon.

[23] See *Wausau Business Insurance Co v Turner Construction Co*, 141 F Supp 2d 412, 417 (SDNY 2001). Enron's bankruptcy is taking place in New York, and a federal court sometimes looks to the law of the forum state. Moreover, many of Enron's transactions with investors took place in New York and would ordinarily be governed by New York law outside of bankruptcy.

[24] See *Harco Energy, Inc v The Re-Entry People, Inc*, 23 SW 3d 389 (Tex App 2000); *Western Horizontal Drilling, Inc v Jonnet Energy Corp*, 11 F 3d 65 (5th Cir 1994). Enron's principal place of business is in Texas.

[25] Substantive consolidation is also used for a purpose altogether different from the issues raised here. Sometimes, the affairs of the two firms may have become so entangled and their assets so meagre that unscrambling the mess may simply not be worth the cost. When the administrative costs of sorting out the obligations of the two firms dwarf the benefits to any group of creditors from keeping the firms separate, it is in everyone's interest to consolidate the two. See *In re The Leslie Fay Companies*, 207 Bankr 764 (Bankr SDNY 1997).

[26] The leading discussion of substantive consolidation can be found in *Union Savings Bank v Augie/Restivo Baking Company, Ltd*, 860 F 2d 515, 518–19 (2d Cir 1988). The DC Circuit put forward a slightly different formulation of substantive consolidation in *Drabkin v Midland-Ross Corp* (*In re Auto-Train Corp*), 810 F 2d 270 (DC Cir 1987), but there is consensus on the doctrine's basic contours.

[27] *In re Bonham*, 229 F 3d 750, 764 (9th Cir 2000).

There is, however, a threshold problem in using substantive consolidation in this context. Substantive consolidation usually involves two estates that are in bankruptcy. In our example, WhiteElephantCo has not filed.[28] Substantive consolidation of a debtor and non-debtor is precedented.[29] Nevertheless, courts are more reluctant to invoke the bankruptcy court's equitable power in such cases, as it involves property that is not yet within its control.

Even if this problem could be surmounted, the creditors are ill-advised to rely too heavily on veil-piercing or substantive consolidation. These actions merge the assets of OfficeCo and WhiteElephantCo, but leave Bank's security interest untouched.[30] Moreover, they do nothing to recover the dividend paid out to the investors.

B. Form, Substance, and True Sales

A firm's transfer of an asset to a special purpose entity is sufficient to remove it from the bankruptcy estate only if the transfer is a 'true sale' that terminates all of the debtor's rights to the asset. OfficeCo may be able to bring Plaza into its bankruptcy estate without invoking theories of veil-piercing or substantive consolidation if it can show that the transfer to WhiteElephantCo was not a 'true sale'.

The standards by which one measures whether a 'true sale' took place are massively undeveloped. Bankruptcy courts will not, of course, accept the parties' characterisation of the transaction. Conventional black letter doctrine suggests that bankruptcy courts should look to substantive non-bankruptcy law to discover whether an asset is 'property of the estate' within the meaning of §541.[31] Whether the transfer of White Elephant Plaza (or more precisely OfficeCo's transfer of its equity interest in the firm that owned White Elephant Plaza) was a 'true sale' would turn on substantive non-bankruptcy law where even the threshold choice of law questions are hard.[32]

[28] Similarly, the SPEs such as LJM2 in Enron have not filed for bankruptcy either.

[29] See *In re Bonham*, 229 F 3d 750, 765 (9th Cir 2000); *Munford, Inc v TOC Retail, Inc*, 115 Bankr 390, 397–98 (Bankr ND Ga 1990).

[30] In Enron, there seem to be a number of transactions in which the investor in the position of Bank failed to take and perfect a security interest in the assets. In these cases, the veil-piercing action would bring about the desired result, but the problem of retrieving assets now in the hands of investors would remain. It seems odd to hold Bank liable, but not the investors.

[31] See *Chicago Board of Trade v Johnson*, 264 US 1 (1924).

[32] Traditional conflicts rules suggest that the forum jurisdiction will look to the conflict rules of the state of incorporation. See *Fletcher v Atex, Inc*, 68 F 3d 1451, 1458 (2d Cir 1995); *Wausau Business Insurance Co v Turner Construction Co*, 141 F Supp 2d 412 (SDNY 2001). But we do see cases in which courts apply the law of the forum. See, eg, *Vuylsteke v Broan*, 17 P 3d 1072, 1074 n 2 (Oregon App 2001).

Some bankruptcy courts have shown a willingness to derive the answer to the question of whether a transfer is a true sale from the bankruptcy court's equitable powers.[33] A debtor's ongoing control over its accounts receivable may be sufficient to make them property of the estate and hence subject the third party to the automatic stay. Using such a test, one could argue that WhiteElephantCo is close enough to OfficeCo that the bankruptcy court should have jurisdiction over the asset. Once OfficeCo is in bankruptcy, CFOPartners is no longer free to sell White Elephant Plaza. OfficeCo may have enough of an interest in the property such that the automatic stay prevents its sale and prevents CFOPartners from taking unilateral action.

Many of the concerns raised about special purpose entities revolve around the question of whether assets transferred to the SPEs are removed from the bankruptcy process. (Indeed, the purpose of many of these transactions is to make the assets in the SPE 'bankruptcy remote.') It can make an important difference whether a firm's property that has been conveyed to the SPE is part of the bankruptcy estate. If it is part of the bankruptcy estate, it can be used in the firm's ongoing operations provided the third party's rights are adequately protected.[34] If the property has been transferred outright, the debtor must reorganise without it. An outright non-recourse sale of accounts is likely a true sale, but other transactions are more problematic. It is not obvious, for example, that a debtor should be able to convey its raw materials to a third party and thus remove them from the bankruptcy estate, if it continues to possess them and process them in its ongoing business.[35]

Existing debates over SPEs, however, are largely useless where a creditor such as Bank has a perfected security interest in the assets. Bank will prevail in the end even if White Elephant Plaza becomes part of the bankruptcy estate. But even if Bank's claim against White Elephant Plaza were unsecured and the investors had not received a dividend, the idea that the sale can be disregarded must be rooted in some established legal doctrine. The body of law that has emerged over whether a transaction is a true sale is quite limited. Bankruptcy law does not establish whether a transaction is a true sale. Instead, bankruptcy looks to non-bankruptcy law. In this case, the relevant law is the law governing security interests. That law, however, allows parties in the position of Bank to prevail as long as they jump through the right hoops. Moreover, states have tried to make it easier to keep assets out of the bankruptcy estate.[36] The transaction here seems quite suspect even

[33] See, eg, *In re LTV Steel Co,* 274 Bankr 278 (Bankr ND Ohio 2001).
[34] See 11 USC §364.
[35] LTV involved a case in which the debtor had conveyed its inventory, including steel in process of being fabricated, to an SPE. The legal basis for disregarding the transaction, however, is unclear, at least if the hurdles of Art 9 have been complied with.
[36] Delaware recently enacted the 'Asset-Backed Securities Facilitation Act.' Without actually defining a 'securitization transaction,' it provides that '[a]ny property, assets, or rights purported

apart from whether the actors complied with whatever formalities were put in their path. Focusing on a true sale therefore, however much it might work under the facts of a particular case, does not squarely focus upon whether transactions of this nature are, as a general matter, suspect. To do this, we must turn to fraudulent conveyance law.

IV. FRAUDULENT CONVEYANCES IN THE POST-MODERN ERA

Fraudulent conveyance law is built on the same principles as substantive consolidation and true sales. Indeed, these doctrines are best understood as being specific elaborations of more general legal principles.[37] Fraudulent conveyance is the most direct way of attacking the various transactions involving OfficeCo and Plaza. Fraudulent conveyance law provides that transfers made and obligations incurred with the intent to 'delay, hinder, or defraud' creditors are fraudulent and void as against creditors. To fall within the reach of fraudulent conveyance law, it is not necessary to prove actual fraud. It is sufficient if the transaction bears 'badges of fraud.'

The most often invoked 'badge of fraud' is a transfer an insolvent debtor makes without reasonably equivalent value. Creditors can avoid an insolvent debtor's birthday present to his mother. The motivations of the debtor are irrelevant. When a debtor is insolvent, he is no longer giving away his own assets. A dividend or a stock repurchase is the corporate analogue to a gift. These transactions are so likely to injure creditors and bring no corresponding benefit to the firm that it makes sense to ban them whenever they leave the debtor insolvent or with unreasonably small capital.

In our example, the $250 dividend to CFOPartners may well have been a fraudulent conveyance. WhiteElephantCo had nothing other than the income the Plaza generated to meet its obligations. Assuming that the Plaza was worth no more than $1000, the firm lacked the resources to repay Bank and OfficeCo at the time it made its dividend. OfficeCo, a creditor of WhiteElephantCo, can seek to set aside the transfer as a fraudulent conveyance. But CFOPartners may be able to argue that White Elephant Plaza was worth more than what they paid for it and that the dividend did not

to be transferred, in whole or in part, in the securitisation transaction shall be deemed to no longer be the property, assets or rights of the transferor.' These provisions were designed to provide a safe harbour to securitization transactions. For example, if this provision were effective in bankruptcy, a firm could establish a Special Purpose Entity into which it transferred its accounts receivable, and investors would be able to acquire interests in the Special Purpose Entity, confident that they would be insulated from any bankruptcy involving the parent. Texas, Alabama and Ohio enacted non-uniform provisions of Revised Art 9 intended to have a similar effect.

[37] See Robert C Clark, 'The Duties of the Corporate Debtor to its Creditors' (1977) 90 *Harvard Law Review* 505.

leave it insolvent. The more volatile the value of the underlying asset, the harder it will be to show insolvency at the moment the dividend is made.[38]

Many suspect transactions that compromise the rights of creditors leave the debtor insolvent and without reasonable equivalent value. We can find, however, other 'badges of fraud' (that similarly require no showing of 'fraud' as traditionally understood) in the case law. Hence, the creditors of OfficeCo do not necessarily need to show insolvency, as long as they can show that other 'badges of fraud' are present.[39] 'Badge of fraud' is a term of art. They identify behaviour that reasonable creditors would prohibit if they could, regardless of whether it would be considered 'fraudulent' at common law.

A transaction is not voidable merely because it contains a single badge of fraud, but transactions that contain a sufficient number of badges of fraud are voidable. 'Badges of fraud' are not clearly defined. Unlike the familiar 'without reasonably equivalent value while insolvent' badge that has taken on a rule-like character, the other badges require traditional common law reasoning and resist easy categorisation. They can be found in part in the case law and in statutes.

The basic ideas are clear. A transfer can be found to 'hinder, delay, or defraud' if it has enough of the following sorts of characteristics:

(1) A close relationship between the parties;[40]
(2) A questionable transfer not in the usual course of business;[41]
(3) The retention of control of the property by the transferor after the conveyance;[42]
(4) The concealment of the transfer.[43]

Fraudulent conveyance law does not require more than that the transfer was part of a plan (claiming earnings were more stable than they were) designed to lull creditors into inactivity. We can take several steps back from OfficeCo's transaction and see that it exhibits many of these traditional 'badges of fraud'.

[38] Recall that in the Raptor III transaction the asset was TNPC stock. Indeed, it was the volatility of the stock price that motivated the transaction in the first place.
[39] In his excellent comment on this paper, Kevin Davis points out that there may be significant costs associated with a fraudulent conveyance law that embraces 'badges of fraud' beyond an insolvent debtor's transfer of assets for less than reasonably equivalent value. In his view, the familiar test picks up the vast majority of suspect transactions, and other doctrines, such as those involving breach of fiduciary duty, pick up many more. See, eg, *Geyer v Ingersoll Publications Co*, 621 A 2d 784, 787 (Del ch 1992). Hence, in Professor Davis's view, the uncertainties of a more expansive fraudulent conveyance regime make it a game not worth the candle.
[40] See *Wall Street Associates v Brodsky*, 684 NYS 2d 244 (App Div 1999).
[41] *Ibid.*
[42] *Ibid.*
[43] UFTA §4(b)(7).

The transaction took the form of a sale of stock. There was no public record of the transfer. The acquiring entity was controlled by an insider of OfficeCo. From the perspective of creditors nothing changed. The day-to-day operations of White Elephant Plaza continued. The owner of record (White Elephant Enterprises) remained unchanged. Debtors, of course, routinely engage in transactions that have, as a side effect, making its balance sheet look healthy and robust. But these transactions offer some collateral benefit, such as favourable tax treatment. The transaction served no purpose other than to make Office Co's earnings appear more stable than they were and to prevent creditors from seeing that the capital spent on Plaza was more at risk than it appeared. Here we have a transaction that unambiguously took resources out of corporate solution and served no purpose other than to make it hard for creditors to understand what its debtor was doing. The absence of a business justification for a transaction has long been a common hallmark of those that are ultimately found to have sufficient 'badges of fraud' to make them fraudulent conveyances.[44]

The OfficeCo transaction involved a number of steps, but these do not matter as fraudulent conveyance law looks to substance rather than form. A debtor cannot escape the reach of fraudulent conveyance law merely by respecting niceties of legal forms and generally accepted accounting principles. The step-transaction doctrine is an essential feature of fraudulent conveyance law.[45] Courts can recharacterise transactions, treat discrete transactions as part of a larger scheme, and otherwise ensure that the substance of the transaction is the focus rather than the form. Indeed, this idea embedded in fraudulent conveyance law is what generates specific doctrines, such as the power of a court to recharacterise as a secured transaction a deal that purports to be an outright sale.

The availability of the fraudulent conveyance remedy can be put more simply. It matters not at all how elaborate the transaction or the nature of the underlying assets. Just as Enron is no different from our hypothetical, our hypothetical would be no different if, instead of equity in a firm that owned an office building, the debtor had transferred sheep to some third party, but continued to possess the sheep, shear them, and mark them as his own. If such a transaction falls within the reach of fraudulent conveyance law even in the absence of actual fraud, then OfficeCo's transaction (and Enron's) should as well. And there is no doubt about the outcome of the case involving sheep. Indeed, the voidability of such a transaction has been one of the foundational principles of Anglo-American law for over 400 years.[46]

[44] See, eg, *Clow v Woods*, 5 Sergeant & Rawle 275 (Pa 1819).
[45] See, eg, *Orr v Kinderhill Corp*, 991 F 2d 31 (2d Cir 1993).
[46] See *Twyne's Case*, 3 Coke 80b, 76 Eng Rep 809 (Star Chamber 1601).

It might seem that Bank's security interest should be respected. It parted with $250 and hence gave value in turn for its security interest. All is not so simple, however. We can recharacterise Bank's transaction as a transfer to CFOPartners for which Bank received a security interest from WhiteElephantCo. Bank's security interest is vulnerable in exactly the same way it would have been had it facilitated a leveraged buyout.[47] Bank will undoubtedly argue that its lien on White Elephant Plaza should be respected at least to the extent that it gave value. The trustee can argue that, in substance, though not in form, the transaction was one in which Bank made a transfer to the investors in CFOPartners and gave nothing to WhiteElephantCo itself. Bank might be able to resist this argument on the ground that it acted in good faith and engaged in an arms' length transaction according to ordinary business terms.

Courts, however, have been quick to question the bona fides of Bank in such a case as this.[48] While a leveraged buyout can serve important aims, a transaction such as this one has no legitimate purpose. In any event, there are few cases in which a fraudulent conveyance is found, but in which courts have nevertheless refused to void the security interest of the party in the position of Bank. Bank's position would be even less tenable if it or an entity related to it also participated as an investor in CFOPartners, as in some of the Enron transactions.

The strength of the fraudulent conveyance attack in cases like Enron may ultimately depend upon the willingness of courts to look to the basic ideas of 'badges of fraud'. The drafters of the Bankruptcy Code and the Uniform Fraudulent Transfer Act slighted their importance. In the vast majority of cases, the badge of fraud that has received explicit codification (transfers by an insolvent for less than reasonably equivalent value) applies with full force. Most suspect transactions within a year of bankruptcy are done by an insolvent debtor for less than reasonably equivalent value. There is no need to look further for other badges of fraud. Similarly, badges of fraud rarely accompany transfers for reasonably equivalent value or fair consideration. Hence, courts too often assume that absent such a transaction actual fraud is required. One can doubt, however, that the bankruptcy court will make this mistake in Enron. New York is one of the few jurisdictions that still has the Uniform Fraudulent Conveyance Act (the predecessor to the modern UFTA). State courts in New York have invoked the idea of 'badges of fraud' even where actual fraud has not been proved and reasonably equivalent value exists.[49] The creditors' lawyers are most unlikely to overlook this body of law.

[47] See *Lippi v Citibank*, 955 F 2d 599 (9th Cir 1992).
[48] See, eg, *United States v Tabor Court Realty Corp*, 803 F 2d 1288 (3d Cir 1986).
[49] See, eg, *Wall Street Associates v Brodsky*, 684 NYS 2d 244 (App Div 1999).

V. CONCLUSION

One can doubt whether we have done enough to craft laws and regulations that prevent the gaming of the system such as we have seen in Enron. Existing legal rules rely too heavily on categories that are rapidly losing significance in the world of commerce. A legal regime that depends upon a fundamental difference between debt and equity is useless in a world in which actors are free to exploit the lessons of Black-Scholes and put-call parity.[50]

By the same measure, legal rules that are focused upon discrete legal entities may prove similarly ineffective. As Coase taught us long ago,[51] transactions that can be done inside a firm can be done outside as well. The only difference lies in the transaction costs associated with the two. In a world in which transaction costs are collapsing both inside the firm and in the market, the difference between the two ways of engaging in the same economic enterprise becomes less important.[52] A computer maker may own a factory or sit in a rented office and control everything with a single laptop.[53] In such a world, preventing those bent upon mischief before the fact becomes increasingly hard.

Nevertheless, the failure of legal rules before the fact should not obscure the ability of the common law principles to right the score after the fact. Those who participated in lucrative deals with Enron will likely suffer the consequences of being too clever by half. Debtors cannot undermine the rights of their creditors by manipulating legal forms and accounting conventions. This was the lesson of *Twyne's Case* in 1601, and it has remained the lode star of bankruptcy law ever since. Those who delight in artifice and contrivance in structuring a debtor's affairs forget it at their peril.

[50] See Fischer Black & Myron Scholes, 'The Pricing of Options and Corporate Liabilities' (1973) 81 *J Pol Econ* 637. For a discussion of how these ideas undermine corporate taxation, see Alvin C Warren, 'Financial Contract Innovation and Income Tax Policy' (1993) 107 *Harvard Law Review* 460.

[51] See RH Coase, 'The Nature of the Firm' (1937) 4 *Economica* 386.

[52] See Douglas G Baird & Robert K Rasmussen, 'The End of Bankruptcy' (2003) 55 *Stan Law Review*.

[53] 'See Incredible Shrinking Plants', *Economist*, 23 Febrauary 2002, at 71.

Commentary on 'Enron and the Long Shadow of Stat. 13 Eliz.': Does the Proper Domain of Fraudulent Conveyance Law Include Deceptive but Fair Transactions?

KEVIN E DAVIS

I. INTRODUCTION

THE ENRON DEBACLE has led many commentators to denounce the American corporate governance regime as an expensive anachronism, chronically incapable of keeping pace with the innovations of sophisticated financiers intent upon gaming the system. Douglas Baird's contribution to this collection bucks that trend. He argues that whatever the failings of other components of the legal system, ancient principles traditionally embodied in the law of fraudulent conveyances can be relied upon to provide creditors with proper recourse against the players who connived with Enron to swindle them out of their money.

In its traditional form, the law of fraudulent conveyances allows a court to set aside transfers made or obligations incurred by a debtor with the intent to 'delay, hinder or defraud' creditors. Most fraudulent conveyances involve transfers at an undervalue, that is to say, transactions in which the debtor receives less than the full value of the property that it has transferred or the obligations it has incurred. However, Professor Baird emphasises the potential breadth of fraudulent conveyance law, noting that in its original form—which has been retained in some jurisdictions—a fraudulent conveyance did not necessarily involve a transaction at an undervalue. He argues that in addition to transactions at an undervalue, fraudulent conveyance law can and should encompass transactions between closely related parties that have no legitimate business justification other than to make it difficult for creditors to understand a debtor's affairs, even if the debtor receives fair value throughout. Professor Baird suggests that it would be not only possible but also positively desirable for the New York bankruptcy court seized with Enron's chapter 11 proceeding to adopt this expansive definition of a fraudulent conveyance. I am not inclined to challenge Professor Baird's interpretation of New York law, but I am inclined to challenge his endorsement of the expansive definition of a fraudulent conveyance.

II. AN EFFICIENCY-BASED APPROACH TO FRAUDULENT CONVEYANCE LAW

For the purposes of this comment I will adopt an approach that Professor Baird helped to pioneer and presume that fraudulent conveyance law ought to be designed to achieve ex ante efficiency, ie, to maximise the net benefits, viewed from an ex ante perspective, that accrue to debtors and those who deal with them.[1] On this view, the law ought to be concerned principally with maximising the size of the pie that a debtor and its creditors can expect to divide among themselves, leaving the division of the pie to be determined through bargaining. This is an attractive approach to lawmaking in a market economy. But this way of describing the efficiency-based approach to debtor-creditor law highlights the difficulty of using it to determine how to regulate transactions that have an adverse impact upon stakeholders who are not in a position to bargain for compensation. However, for reasons that will be outlined below, I believe that this point has limited bearing on the matters discussed in this comment.

The central implication of adopting efficiency as a guiding principle in the formulation of fraudulent conveyance law is that the transactions vulnerable to attack should be confined to those transactions that are unlikely to yield net benefits for the parties. A transfer at an undervalue that renders the transferor insolvent is the paradigmatic example of such a transaction. To see this, suppose that C1 lends $100 to D1 who has no other assets. D1 then gratuitously transfers the $100 to T1, thereby rendering itself incapable of satisfying its debt to C1. This transaction is likely to yield a net loss to the parties: C1 has lost $100 and even if we assume that D1 has access to some or all of the value of the transferred assets in the hands of T1, D1's benefit is unlikely to exceed $100. In fact, once the costs of the transaction are taken into account D1's benefit may well be less than $100 and so the net benefits to the parties will be negative.

The difficulty these parties face is that ex post D1's residual claimants (eg, its shareholders if D1 is a corporation) have a strong incentive to engage

[1] DG Baird and TH Jackson, 'Fraudulent Conveyance Law and its Proper Domain' (1985) 38 *Vand Law Review* 829. See also, TH Jackson, 'Avoiding Powers in Bankruptcy' (1984) 36 *Stan Law Review* 725 at 777–86.

in this sort of transaction. In the absence of the transaction D1's assets will all be used to satisfy C1's claim, generating no benefit for D1's residual claimants. However, if those residual claimants have any sort of claim, whether legal or moral, against T1, then they may derive at least some benefit from the assets in the hands of T1—hence the incentive to transfer D1's assets to T1. In fact, generally speaking, whenever the value of the fixed claims against a debtor equals or exceeds the value of its assets its residual claimants have an incentive to transfer its assets at an undervalue to a related party, even if the transaction costs of doing so are significant.

The situation is quite different however, when viewed from an ex ante perspective, that is to say, if we consider the parties' views prior to C1's extension of credit to D1. Ex ante the parties would find it mutually beneficial to find some way of discouraging D1 and T1 from engaging in transactions that yield negative net benefits. Fraudulent conveyance law is one of several legal doctrines that can serve this purpose.

Professor Baird's paper provides an important reminder that transfers at an undervalue in the vicinity of insolvency are not the only types of transactions that can yield negative net benefits for debtors and their creditors. In fact, I can think of at least one other category of transactions that is likely to fit this definition. Consider a scenario in which a debtor sells an asset that is crucial to the operation of its enterprise to a related party at fair market value. A good example might be a newspaper company ('Newsco') selling its custom-built—and thus difficult to replace—printing press to an affiliate ('Relatedco'). The fair market value of the printing press may not reflect its value to Newsco's creditors. This is because Newsco's other assets will be of little value to the creditors unless they can also obtain access to the printing press, but in attempting to obtain such access they will be exposed to the risk of opportunistic behaviour on the part of Relatedco. Newsco's current management may, however, reasonably discount the possibility of being exposed to such opportunism. Thus, even if the transaction significantly diminishes the value of Newsco's assets to its creditors, it will not necessarily cause a corresponding reduction in the value of Newsco to its shareholders.[2] Such a transaction will be inefficient if it serves to diminish the welfare of Newsco's creditors without, once the costs of

[2] *Agricultural Mortgage Corporation v Woodward* [1995] 1 BCLC 1 (CA), may provide another example of a transaction of this sort. In that case a farmer who was facing foreclosure granted a tenancy in his farm to his wife at what was conceded to be a full market rent. The grant of the tenancy was set aside as a transaction at an undervalue. However, this characterization is difficult to reconcile with the concession that the rent was a fair one. The real concern may have been that the debtor substituted a farm subject to a tenancy for a farm capable of being sold with vacant possession, with the former being much more valuable to the debtor's creditors—given their intended disposition of the property—than the latter. The diminution in the value of the property to the creditors may not, however, have been matched by a diminution in its value to the farmer since granting the tenancy to his wife probably did not interfere with the farmer's intended use of the property.

effecting the transaction are taken into account, creating a fully offsetting enhancement in the welfare of Newsco's other stakeholders. However, the fact that such a transaction involves, by stipulation, the sale of an asset at fair market value means that it is difficult to characterise as a transfer at an undervalue.[3]

Of course in practice transactions do not arrive before counterparties (or courts) neatly labelled as being either efficient or inefficient. Often it is quite difficult to distinguish inefficient transactions from efficient ones. For instance, the fair market value of property transferred by a debtor may be difficult to ascertain, making it difficult to determine whether the property is being transferred at an undervalue. In addition, even in cases involving transfers at an undervalue it is crucial to challenge only transactions that are highly likely to render the debtor insolvent. Other transfers at an undervalue may have positive effects on managerial incentives that eliminate or offset any adverse effects on creditors. Similarly, transactions that involve transferring critical assets to related parties may generate offsetting tax benefits or make it easier to monitor the use of the assets in question.

The difficulty of distinguishing efficient from inefficient transactions leads us to a second implication of using efficiency as a guide when determining the scope of fraudulent conveyance law: it is important to consider the costs that fraudulent conveyance law imposes upon a debtor's potential counterparties. These costs generally comprise the cost of the precautions that potential counterparties will take in order to minimise the risk of having a transaction set aside as a fraudulent conveyance together with the cost to those parties of bearing any residual risk. For example, faced with the prospect of liability under fraudulent conveyance law, counterparties may insist upon receiving an independent valuation of property being transferred or upon conducting an in-depth analysis of whether there is any legitimate business justification for the transaction. Alternatively, some parties may avoid an entire category of transactions altogether. All of these precautions are costly. Consequently, an efficient fraudulent conveyance law will ensure that the costs of recruiting a debtor's prospective counterparties to assist in preventing any given type of inefficient transaction do not outweigh the benefits.

Adopting an efficiency-based approach to the analysis of fraudulent conveyance law has a rich set of legal implications. For instance, it suggests

[3] See, however, *Agricultural Mortgage Corporation* above in which the Court of Appeal suggested that an asset's 'surrender value' must be taken into account in assessing its fair market value (See also, R J Mokal and L C Ho, 'Consideration, Characterisation, Evaluation: Transactions at an Undervalue after *Phillips v Brewin*' [2001] *JCLS* 359). The surrender value of an asset originally held by a debtor appears to equal the difference between its value to a debtor's creditors and the value to those creditors of any assets received in substitution. In other words, surrender value is a measure of the costs that a transfer of an asset imposes upon the transferor's creditors. If the benefit that a debtor obtains by transferring an asset exceeds the asset's surrender value then the transaction will be efficient in the sense used in this comment.

that the legislative provisions that form the basis of the modern English law of fraudulent conveyances may be under-inclusive as a result of the fact that they only capture transactions at an undervalue and do not clearly include, for example, fair market value dispositions of critical assets.[4] It also suggests that courts should be cautious about setting aside transactions involving parties who did not know and could not reasonably have known that they were participating in a transaction at an undervalue.[5] Many of these implications have been explored elsewhere by other scholars.[6] Thus, in the remainder of this comment I will focus upon the implications of this mode of analysis for Professor Baird's proposed interpretation of New York fraudulent conveyance law.

III. SHOULD FRAUDULENT CONVEYANCE LAW CAPTURE TRANSACTIONS THAT ARE SIMPLY DECEPTIVE?

Professor Baird suggests that fraudulent conveyance law should capture transactions designed to make it difficult for creditors to understand a debtor's affairs. Based upon the examples he provides, here Baird seems to have in mind transactions that involve transferring assets to a related entity in return for some sort of financial claim against the transferee. A simple example involves D2 transferring assets to T2 in exchange for a promissory note. Unless the value of the promissory note is constantly updated and reported to D2's creditors, those creditors will arguably find it more difficult to ascertain the value of D2's assets than if the transaction had never taken place.

The difficulty with Professor Baird's proposal is that it does not clearly comport with either the idea that law of fraudulent conveyances should protect the interests of involuntary creditors or the view that the law should be designed to discourage inefficient transactions. As far as the first idea is concerned, it is important to recognise that the impact of deceptive transactions only falls, at least in the first instance, upon creditors who are interested in attempting to assess the value of their debtor's assets. This means that these transactions will typically only impose costs upon relatively sophisticated voluntary creditors, as other creditors are not in a position to

[4] Insolvency Act 1986, ss 238, 423.

[5] See, for example, *Re Taylor Sinclair Capital Ltd* [2001] 2 BCLC 176. Cf Mokal and Ho, above, who criticise this decision on the basis that '… the purpose of section 238 is 'to restore to the insolvent (for the benefit of its creditors) excess value transferred to another …'. Mokal and Ho's premise is not necessarily consistent with the view that only inefficient transactions ought to be set aside as it would prohibit transactions that cause net harm to creditors where the harm is outweighed by the benefit to another party. Perhaps more significantly, Mokal and Ho's approach appears to ignore any costs that might be imposed upon potential counterparties.

[6] See, for example, Baird & Jackson above; J Armour, 'Transactions at an Undervalue' and 'Transactions Defrauding Creditors' in H Bennett and J Armour (eds), *Vulnerable Transactions in Corporate Insolvency* (Oxford, Hart Publishing, 2003), chs 2 and 3.

monitor their debtors' financial situations. This implies that the real benefit of Professor Baird's proposal lies in its potential to influence debtors' behaviour towards relatively sophisticated voluntary creditors.

As suggested above, where the principal interests at stake are those of voluntary creditors and their debtors ex ante efficiency is an attractive guiding principle. There is no doubt that deceptive transactions impose costs upon creditors and so are potentially inefficient. However, the potential benefits of these transactions also need to be considered before any final conclusions can be drawn. As indicated above, the transactions of greatest concern are likely to be transfers of non-financial assets or simple financial assets to related parties in exchange for complex financial assets. Sometimes the benefits of these sorts of transactions will flow from allocating assets with different risk profiles to different legal entities so as to make it easier to monitor the performance of each entity's managers or to make the entities attractive to different types of investors. Alternatively, the benefits may arrive in the form of tax savings associated with having one entity rather than another hold assets. Before it is possible to determine whether any given type of transaction is inherently inefficient these potential benefits must be set against the costs incurred by creditors whose ability to monitor their debtors is impeded. For many types of transactions this balancing exercise may generate indeterminate results.

Suppose, however, that we assume that it is possible to delineate with reasonable clarity a set of transactions that are designed exclusively to exploit loopholes in applicable accounting rules with a view to misleading creditors. It seems reasonable to assume that transactions falling into this category are inherently inefficient. However, it does not necessarily follow that it is efficient to use fraudulent conveyance law to discourage these sorts of transactions. That is because targeting this class of transactions will inevitably impose costs upon transferees. Applying fraudulent conveyance law to transactions designed solely to mislead creditors will be inefficient if this tactic causes debtors' counterparties to incur costs that exceed the benefits of recruiting their assistance in preventing deception.

In assessing the costs that Professor Baird's proposal would impose upon counterparties it is important recognize that the proposal only involves using fraudulent conveyance legislation to target transactions in which the counterparty has not acted in good faith. The fact that an absence of good faith would be a precondition to liability limits the costs that Professor Baird's proposal would impose upon potential counterparties. It would not, however, eliminate those costs because the concept of good faith is highly imprecise and seems to include absence of constructive notice of misconduct.[7] In other words, acting in good faith seems to involve undertaking a certain amount of investigation. Therefore, Professor Baird's

[7] See generally, JF Williams, 'Revisiting the Proper Limits of Fraudulent Transfer Law' 8 *Bankr Dev J* 55 at 108–109.

proposal seems likely to encourage many parties either to scrutinize their trading partners' financial reporting or to categorically reject entire categories of transactions, including both efficient and inefficient ones.

Ultimately, the issue of whether the costs of encouraging counterparties to take precautions against participating in deceptive transactions outweigh the benefits can only be resolved on empirical grounds. However, there are strong theoretical reasons to believe that this strategy would be inefficient. The crux of the matter is that a debtor's counterparties are, relatively speaking, poorly situated to identify transactions designed to mislead other parties. Determining whether a transaction is likely to deceive involves ascertaining not only the terms of the transaction, but also, the way in which it is described to creditors. This means that if a firm's counterparties are to become involved in blowing the whistle on deceptive transactions they will not be able to rest after undertaking a careful scrutiny of the terms of their transactions with the firm. They will also have to wait until those transactions have been reported to the firm's creditors—typically in the form of financial statements that disclose only the aggregate effects of a large number of transactions—and scrutinise those reports to ensure that they accurately reflect the prior dealings. In essence, this regime would require a firm's counterparties to assume the role that its auditors and outside directors have traditionally played. Given the number of other well-informed parties capable of monitoring the accuracy of firms' communications with their creditors, the marginal benefits of enlisting counterparties in this exercise seem small, and unlikely to outweigh the costs.[8]

Of course, events such as Enron's failure suggest that the traditional set of gatekeepers can sometimes fail spectacularly to fulfil their responsibilities. However, these events do not directly undermine the point being made here. My claim is not that the performance of traditional gatekeepers cannot be improved upon, but rather that the costs of enhancing the performance of traditional gatekeepers—eg, through tighter regulation of conflicts of interest—are likely to be lower than the costs of recruiting an additional set of gatekeepers. I should also add that if Professor Baird's positive claim about the content of US fraudulent conveyance law is correct then Enron's counterparties should already have been on notice of the possibility that they could be held liable under fraudulent conveyance law for facilitating deceptive transactions. Nevertheless, according to Professor Baird, neither those counterparties nor Enron were deterred. Thus, ironically, the Enron saga arguably demonstrates the shortcomings of a strategy that relies upon counterparties to discourage transactions designed to mislead creditors.

[8] It is important to emphasize that the analysis in the text is concerned only with counterparties' ability to participate in controlling transactions that are objectionable solely because of the manner in which they are subsequently represented. Counterparties may be relatively well placed to aid in controlling transactions that are more inherently problematic, such as those that involve a clear and immediate breach of fiduciary duty.

IV. ENRON AND TWYNE'S CASE REVISITED

Professor Baird's argument derives considerable rhetorical force from being presented as being consistent with an ancient precedent yet still offering a solution to the challenges facing Enron's creditors. However, I would argue that there is no need to resort to his expansive interpretation of a fraudulent conveyance in order to deal with either Enron or its seventeenth century precursor.

Let us consider the application to Enron first. If we assume that Enron's actual transactions are substantially similar to the hypothetical transaction that Professor Baird uses to illustrate his analysis there is no need to embrace an expansive interpretation of fraudulent conveyance law in order to provide recourse to Enron's creditors. Professor Baird's illustration essentially involves D3 transferring assets to T3 in exchange for a combination of cash and a promissory note. T3 then obtains a loan from T4 secured against its newly acquired assets and distributes the proceeds to the owners of T3. In this scenario, the distribution to the owners of T3 is clearly a transfer at an undervalue which might be vulnerable to attack by T3's creditors, and, indirectly, by the creditors of those creditors. There is no need to resort to the proposition that a transaction can be set aside simply on the grounds that it is deceptive.

As for *Twyne's Case*,[9] Professor Baird is correct to point out that it may not have involved a transfer at an undervalue. In fact, *Twyne's Case* is probably best understood as an example of a preference.[10] However, I believe that the transaction at issue in that case probably did, when it took place, fall within the bounds of an efficiently designed fraudulent conveyance law. In *Twyne's Case*, reduced to its essential elements, D4 purportedly transferred property to T4 in satisfaction of a pre-existing debt but was allowed to retain possession and use of the transferred property. In a world in which creditors probably relied heavily upon ostensible ownership of property to gauge the financial condition of their debtors this transaction was inherently misleading. In fact, it is quite possible that in that context, the costs of the confusion resulting from behaviour of this sort of transaction outweighed the benefits. Moreover, these consequences would have been immediately and readily obvious to T4 since the essence of the transaction involved separation of ownership and possession. Consequently, at the time that *Twyne's Case* was decided, it was probably reasonable to argue both that the type of transaction at issue was inherently inefficient and that counterparties to such transactions could identify them at minimal cost. These arguments lead to the conclusion that the transaction was an appropriate

[9] 3 Coke 80b, 76 ER 809 (Star Chamber, 1601).
[10] RC Clark, 'The Duties of the Corporate Debtor to its Creditors' (1997) 90 *Harvard Law Review* 505.

target for avoidance under the Statute of Elizabeth. But they do not lead to the conclusion that a similar transaction would be an appropriate target for modern fraudulent conveyance legislation. In modern times existing and prospective creditors assess the financial condition of their debtors by reference to public registries and financial statements rather than ostensible ownership and so transactions that lead to separation of ownership and possession are not inherently misleading.[11]

V. CONCLUSION

At first blush Professor Baird's claim that neglected aspects of traditional legal principles may offer solutions to the most quintessentially modern of corporate governance problems is extremely appealing. To the extent that he refers to the traditional principles concerning transactions at an undervalue few would challenge his claim. However, Professor Baird's argument in favour of the more controversial elements of the traditional approach to fraudulent conveyances is unconvincing. Even the best established of legal principles requires updating from time to time.

[11] This appears to be the basis upon which Lord Blackburn argued that *Twyne's Case* was superseded in the nineteenth century by the provisions of the Bills of Sales Acts concerning registration of interests in personal property. See, *Cookson v Swire* (1884) 9 AC 653 (HL), 664–665.

Part 6

Moving Forward: Law and Practice

21

Commercial Law and the Limits of the Black Letter Approach

ANTHONY J DUGGAN*

For the rational study of law the black letter man may be the man of the present, but the man of the future is the man of statistics and the master of economics.[1]

I. INTRODUCTION

SIR ANTHONY MASON, the former Australian Chief Justice, has argued that law and economics scholarship is not much use to judges.[2] He gives three main reasons. First, he says, economic analysis has no contribution to make when the legal issues are non-economic ones.[3] Secondly, contest and debate about economic issues would add to the length and cost of litigation.[4] And thirdly, 'if counsel present an argument based on economic analysis which suggests that judgment for the defendant would lead to wealth maximisation for society, how does a court take account of this if previous authorities or considerations of justice or morality point in the other direction?'[5] Mason's focus is on law and economics, but his case is really against instrumentalism at large and in favour of the black letter approach.[6] Judges are not supposed to think in an 'instrumental'

* My thanks to Ed Miller for research assistance and Michael Trebilcock, Stephen Waddams, Kevin Davis and Megan Richardson for helpful comments on an earlier draft. All errors are mine.
[1] Oliver Wendell Holmes, 'The Path of the Law' (1897) 10 *Harvard Law Review* 457, quoted in Hon Sir Ivor Richardson, 'Law and Economics—And Why New Zealand Needs It' (2002) 8 *New Zealand Business Law Quarterly* 151 at p 156.
[2] 'Law and Economics' (1991) 17 *Monash University Law Review* 167.
[3] *Ibid*, p 172.
[4] *Ibid*, p 174.
[5] *Ibid*, p 181.
[6] Michael J Trebilcock, 'The Value and Limits of Law and Economics' in Megan Richardson and Gillian Hadfield (eds), *The Second Wave of Law and Economics* (Sydney, Federation Press, 1999), 12 at p 23.

way because they have no mandate to prefer one social value over another.[7] Instead, they should decide cases 'within the framework of traditional legal principles', by implication because that way they remain value neutral.[8]

Mason rejects the law and economics approach as a basis for judicial decision-making and, by extension, the practice of law more generally. His argument is open to criticism on the following grounds.

(1) It assumes that legal issues are more often than not non-economic ones and that economics has nothing useful to say in such cases. This overlooks the close inter-connectedness between law and economics. Economics is the science of rational choice. It is 'the study of rational behaviour in the face of scarcity. If there were an abundance of every good thing, there would be no need for law, no need for a state'.[9] Choice is central to most legal disputes. The negligent tortfeasor chooses not to take precautions against harm. The contractor chooses to deal. The criminal chooses to commit the crime. Whenever there is a question of choice, economics has something to say that will be relevant to the understanding of existing laws and the shaping of new ones.

(2) Mason assumes that if the courts adopted an economic approach, they would need outside experts to help them.[10] This concern over-estimates the complexity of the economic issues that most cases involve.[11] The basic point is a very simple one, namely that all legal rules have consequences and consequences matter. 'Every time a court starts looking at the likely effects of its decision, it is engaged in instrumental analysis'.[12] *Barclays Bank plc v O'Brien*,[13] the spousal guarantees case, is a good example.[14] In *O'Brien's* case, Lord Browne-Wilkinson identified the following policy considerations:[15]

[7] Mason, *op cit* p 181.
[8] *Ibid*, p 180.
[9] Frank H Easterbrook, 'The Inevitability of Law and Economics' (1989) 1 *Legal Education Review* 3.
[10] See text at n 3, above.
[11] See Richardson, *op cit* p 172: 'I do not believe that judges are incapable of understanding the relatively simple economic concepts required for most law and economic analyses'. (Richardson is a former President of the New Zealand Court of Appeal and a contemporary of Mason's.)
[12] Easterbrook, *op cit* p 4.
[13] [1994] 1 AC 180.
[14] See Richardson, *op cit, passim,* for other examples. Richardson (at p 164) contrasts *Stovin v Wise* [1996] AC 923 with *Dietrich v R* (1992) 177 CLR 292. In *Stovin v Wise*, Lord Hoffmann emphasised that before extending the duty of care owed by public authorities, the courts must consider the cost to the community of the defensive measures they are likely to take to avoid liability. On the other hand, in *Dietrich v R*, the Australian High Court held that the courts should stay criminal proceedings where defendants cannot afford representation and the cost is not borne by the State. The consequence of the decision, which the court failed to address, was to put budgetary pressures on Australian State governments.
[15] [1994] 1 AC 180 at p 188.

> Society's recognition of the equality of the sexes has led to a rejection of the concept that the wife is subservient to the husband in the management of the family's finances ... yet ... although the concept of the ignorant wife leaving all financial decisions to the husband is outmoded, the practice does not yet coincide with the ideal ... In a substantial proportion of marriages, it is still the husband who has the business experience and the wife is willing to follow his advice without bringing a truly independent mind and will to bear on financial transactions. The number of recent cases in this field shows that in practice many wives are still subjected to, and yield to, undue influence by their husbands.

and

> It is easy to allow sympathy for a wife who is threatened with the loss of her home at the suit of a rich bank to obscure an important public interest, viz the need to ensure that the wealth currently tied up in the matrimonial home does not become economically sterile. If the rights secured to wives by the law render vulnerable loans granted on the security of matrimonial homes, institutions will be unwilling to accept such security, thereby reducing the flow of loan capital to business enterprises.

This is law and economics pure and simple.[16] The first passage identifies the benefits of a rule to protect the wife. The second passage identifies the costs. The judgment goes on to fashion a rule that maximises the net benefits.

(3) Mason confuses positive and normative economic analysis: 'in formulating and applying principles, judges take account of many considerations such as precedent and history, as well as morality, culpability, justice and fairness and do not regard themselves as being at liberty to subordinate these considerations to the dictates of economic goals'.[17] True enough, but it is one thing to say that a judge should not be a slave to economic considerations. It is quite another thing to say that a judge should ignore economic considerations altogether. As Trebilcock points out, Mason's view of judicial lawmaking is both atheoretical and non-empirical.[18] Courts that steadfastly refuse to look in a theoretical or empirical way at the consequences of their decisions are reduced to 'muddling through'.[19]

[16] See Michael J Trebilcock and Steven B Elliott, 'The Scope and Limits of Legal Paternalism: Altruism and Coercion in Family Financial Arrangements' in Peter Benson (ed), *The Theory of Contract Law* (Cambridge UP, 2001), p 45.
[17] *Op cit* p 174.
[18] *Loc cit.*
[19] *Ibid*. The 'muddling through' reference comes from C Lindblom, 'The Science of Muddling Through' (1959) 19 *Public Administration Review* 79.

The theme of this seminar is a suggested misalignment between commercial law as expounded and professed by judges and academics and the commercial expectations of the business community. If there is such a gap, it is most likely the product of a muddling through strategy. To muddle through is to 'move on certain margins in a tentative fashion, operating largely by trial and error, in trying to find a solution'.[20] Muddling through is hit and miss. In the commercial law context, the business community will applaud the hits and condemn the misses. The predominant attitude in business circles to commercial law at large is likely to vary according to the ratio of hits to misses from time to time. The way to close the gap is to improve the ratio. To improve the ratio, courts must abandon the muddling through strategy, to the extent that they still practise it. In other words, they must address the economic consequences of their decisions. Sir Ivor Richardson makes this point forcefully in a paper that contrasts sharply with the Mason lecture. In Richardson's words, 'as judges, and inevitably as lawmakers, we must not make laws without regard to their full costs and benefits to the community'.[21] In other words, muddling through is no longer good enough.

The balance of this paper discusses three commercial law topics: (1) fiduciary obligations in commercial relationships; (2) reservation of title clauses in sales contracts ('Romalpa agreements'); and (3) equitable remedies in bankruptcy. The aim in each case is to test the limits of the black letter approach and to demonstrate the insights a law and economics perspective can offer.

II. FIDUCIARY OBLIGATIONS IN COMMERCIAL RELATIONSHIPS

In *Pilmer v The Duke Group (in liq)*,[22] a recent decision of the Australian High Court, the question was whether financial advisers owe fiduciary duties to their clients. Kia Ora Gold Corp NL ('Kia Ora') was a public listed company. It made a takeover offer for Western United Ltd. The takeover price was a combination of cash and Kia Ora shares. In total, Kia Ora paid out $25.7 million cash to Western United shareholders and it allotted 67.9 million $1 shares. The October 1989 stock market crash occurred while the takeover was in progress. Western United's share price dropped dramatically and Kia Ora ended up making a substantial loss on the deal. The Australian stock exchange listing rules prohibit a listed company from buying shares without the prior approval of its shareholders in general meeting. The rules say that a report must be sent to shareholders with the notice of meeting. The report, to be prepared by independent

[20] Trebilcock, *loc cit*.
[21] *Op cit* p 173.
[22] (2001) 75 ALJR 1067, noted (2002) 118 LQR 180.

qualified persons, must establish that the takeover price is a fair one. The Kia Ora board hired Nelson Wheeler, a firm of accountants, to prepare the report. Nelson Wheeler put in a report saying that the takeover price was fair and reasonable. As it happened, the report was incompetently prepared. Nelson Wheeler should have seen that Kia Ora's takeover bid substantially overvalued Western United.

Kia Ora went into liquidation. The liquidator sued the former Kia Ora directors and Nelson Wheeler for recovery of Kia Ora's losses on the takeover deal. He sued Nelson Wheeler in contract and tort, and also for breach of fiduciary duty. The liquidator argued that Kia Ora suffered a loss when it issued the new shares and allotted them to the Western United shareholders. The alleged loss was the opportunity cost of the capital the issued shares raised. The fiduciary claim was added in the hope that the liberal rules governing assessment of equitable compensation might cover this loss even if the contract and tort rules did not. In support of the fiduciary claim, the liquidator argued that: (1) the financial adviser-client relationship is a fiduciary one; (2) the fiduciary's duties include a duty to avoid conflicts of interest; and (3) Nelson Wheeler had a conflict of interest because of their ongoing association with the Kia Ora board. There was a conflict of interest because they owed a duty to Kia Ora to report impartially on the takeover offer price, but at the same time they had a vested interest in reporting favourably so they would get more work from Kia Ora in future. To avoid the conflict of interest, Nelson Wheeler should have declined the retainer.

The majority of the court (McHugh, Gummow, Hayne and Callinan JJ) dismissed the claim for breach of fiduciary duty. It said there was no real or substantial possibility of any conflict of interest. Proof of past dealings between a financial adviser and a corporate client does not itself establish a conflict of interest. Nor is it enough to say generally that the financial adviser had a hope or expectation of future dealings. 'That will often be so. Most professional advisers would hope that the proper performance of the task at hand will lead the client to retain them again'.[23] Kirby J dissented. He said that 'the company relied on [Nelson Wheeler] for an independent, impartial and competent report which [Nelson Wheeler] were incapable of providing, and did not provide'.[24] 'In circumstances where, in effect, undivided loyalty to Kia Ora and its unsuspecting shareholders was impossible, had not been, and could probably not have been, repaired by disclosure, the only way in which Nelson Wheeler could have discharged their fiduciary obligations was by declining to act'.[25]

A recent Law Quarterly Review note suggests that there is nothing to choose between the competing views in the *Pilmer* case on the breach of

[23] Para 83.
[24] Para 137.
[25] Para 142.

fiduciary duty question: 'different minds may reach different conclusions as to the presence or absence of a real or substantial possibility of conflict between duty and interest or between duty and duty'.[26] Therefore, 'there is little, if anything, a commentator can say that does not amount simply to unhelpful second-guessing of the facts in a case'.[27] This is black letter scholarship giving up the ghost. What does the law and economics approach have to say?

Kirby J's judgment has implications that extend beyond the particular facts of the case. It suggests there may be a conflict of interest in any case where an accounting firm has an ongoing relationship with the company's board. The logical extension is that a company could not safely use the same auditors from one year to the next. There would always be the risk of one year's report being coloured by the firm's expectation of re-engagement for the next year's audit. If this were the law, the consequence would be to increase substantially the cost to companies of audits and financial consulting services. The advantage of using the same auditor from one year to the next is that the auditor can achieve economies of scale over time. The firm acquires background information about the company first time round. Second time round it avoids the cost of having to acquire the information again. The firm presumably passes on the savings so that the company and its shareholders benefit. These savings would be lost if companies had to replace their auditors on a regular basis.

This is the main cost of Kirby J's approach, at least as it applies by extension to auditors. The benefit is fewer conflicts of interest and so better protection for investors. Does the benefit exceed the cost? Probably not, because in the usual case there are sufficient incentives anyway for a company's board not to fire its auditors for failing to toe the management line.[28] 'Firing is highly salient'.[29] A company that fires its auditor will attract publicity and people will want to know the reasons why. The fired auditor itself may well go public on the issue, but even if it remains silent news of the event will cause questions to be asked.[30]

These considerations support the majority's view that proof of an ongoing relationship with the company's board is not itself enough to disqualify the financial adviser. However, that is not the end of the story. The majority judgment suggests that additional facts may tilt the balance in favour of judicial intervention. It does not say what kind of additional facts might make the difference. The Enron scandal offers some clues.

[26] (2002) 118 LQR 180 at p 183, quoting from the majority judgment in the *Pilmer* case at para 79.
[27] (2002) 118 LQR 180 at p 183.
[28] But see text at n 31 and following, below.
[29] Jeffrey N Gordon, 'What Enron Means for the Management and Control of the Modern Business Corporation: Some Initial Reflections' (2002) 69 *University of Chicago Law Review* 1233 at p 1237.
[30] *Ibid*.

Enron's accountants, Arthur Andersen, certified falsely that the company's financial statements 'fairly represented' its financial picture. This looks like *Pilmer* writ large. The difference is that not only did Arthur Andersen have an ongoing relationship with the Enron board, it also cross-sold its consulting services to the company. Gordon explains why this difference matters:[31]

> [the] picture dramatically changes when the accounting firm begins to cross-sell consulting services. It is not that the accountant now has more at stake in the relationship and thus would lose more if fired by the company. Nor is it simply that the accountant may now have a particular reason to please, or at least not alienate, the client who may buy additional services, and may even hope that cooperation on difficult accounting questions will be appreciated as part of a total client relationship. Rather, it is that the client now has available a repertoire of 'low visibility sanctions' to discipline the accountant's behaviour. If the accountant is resistant, a contract may be withheld or not renewed (or if the accountant is cooperative, the reverse). But unlike the firing, these disciplinary measures will not be disclosed ... Thus the issuer now has credible threats against an accountant who disagrees with management on an important issue.

As Gordon goes on to say, the Enron scandal challenges the efficient market hypothesis. Arthur Andersen's lack of independence from Enron was known to sophisticated investors, as was the weakness of Arthur Andersen's internal governance mechanisms in controlling partners' behaviour. This information should have been reflected in Enron's stock price from the outset, but apparently it was not.[32] One solution to this apparent market failure lies in fiduciary law. It can plausibly be said that the cross-selling of consulting services by an auditor gives rise to a conflict of interest for the reason Gordon suggests. To resolve the conflict, the firm should have to give up either its audit retainer or the consultancy, on pain of suffering the consequences in equity. The United States has opted for a statutory solution to the problem. The Securities Exchange Act, as recently amended, prohibits a corporation's auditor from performing specified non-audit services for the corporation, including bookkeeping services, financial information systems design and implementation, appraisal or valuation services, fairness opinions and management functions or human resources.[33] An auditor may engage in any non-audit service that is not on the statutory list of prohibited services, including tax services, but only if the corporation's audit committee approves the activity in advance.[34]

[31] *Ibid*, pp 1237–1238.
[32] *Ibid*, pp 1238–1239 and 1240.
[33] Securities Exchange Act of 1934, s 10A(g), added by the Sarbanes-Oxley Act of 2002, s 201.
[34] *Ibid*, s 10A(h).

III. RESERVATION OF TITLE CLAUSES

Associated Alloys Pty Limited v CAN 001 452 106 Pty Limited (in liquidation)[35] is a recent decision of the Australian High Court dealing with the application of the registration of charges provisions in what was then Corporations Law, Part 3.5 (now chapter 2K) to the proceeds clause in a conditional sale agreement. The Corporations Law requires the registration of a charge on designated property of a company, including book debts. 'Charge' is defined to mean 'a charge created in any way and includes a mortgage and an agreement to give a charge or mortgage, whether on demand or otherwise'. Failure to register means that the charge is void as against a liquidator or administrator of the company. The appellant ('Seller') sold steel to the respondent ('Buyer') under a series of conditional sale agreements made between 1981 and 1996. The steel was for use in Buyer's manufacturing processes. Seller's invoices contained a reservation of title clause which said that title to the goods was to remain in Seller until Buyer paid the price in full. The reservation of title clause contained a proceeds subclause, which read as follows:

> In the event that [Buyer] uses the goods/product in some manufacturing or construction process of its own or some third party, then [Buyer] shall hold such part of the proceeds of such manufacturing or construction process as relates to the goods/product in trust for [Seller]. Such part shall be deemed to equal in dollar terms the amount owing by [Buyer] to [Seller] at the time of the receipt of such proceeds.

Buyer went into liquidation in 1996. Seller had not been fully paid and it asserted its rights under the proceeds subclause against Buyer's liquidator. A conditional sale agreement is not itself a registrable charge within the meaning of the Corporations Law. However, Buyer's liquidator argued that the proceeds subclause created a charge on Buyer's book debts and it was void for non-registration. Seller argued that the proceeds subclause created a trust in its favour, not a charge, and a trust is not registrable. Accordingly, the case turned on the difference in equity between a charge and a trust. The trial judge held that, despite the parties' use of trusts language, the proceeds subclause created a charge. The New South Wales Court of Appeal agreed.

The High Court, by a majority, did not. It saw the key question as being whether the parties intended to create a trust and it treated their explicit use of trusts language as proof that they did. 'There is nothing to suggest', the court said, 'that the parties in their written instrument did not mean what they said, or did not say what they meant'.[36] Therefore, there was no

[35] (2000) 171 ALR 568.
[36] (2000) 171 ALR 568 at para 35.

reason for not taking the parties at their word or for giving the bargain a different meaning from the way the parties themselves had expressed it.[37] The court recognised the commercial significance of its decision, saying that 'for third parties, such as financial institutions seeking to assess the credit-worthiness of the buyer, the non-registration of the Proceeds Subclause on a public register may cause potential difficulties'.[38] However, it suggested that these difficulties were capable of remedy by legislation'.[39] 'To treat the Proceeds Subclause as an agreement which falls foul of the Law is to rewrite the statute. It is not for the courts to destroy or impair property rights, such as those arising under trusts, by supplementing the list of those rights which the legislature has selected for such treatment'.[40] Kirby J, in a vigorous dissent, cautioned against the 'exclusive concentration on the terms of an instrument purporting to create a trust, to the neglect of an examination of the purpose and effect of the instrument when considered for its substance not merely its form'.[41] He concluded that the proceeds subclause did create a registrable charge and he said that any other construction 'would permit the easy defeat of the clear purpose of the Law'.[42]

Article 9 of the United States Uniform Commercial Code governs security interests in personal property. It applies to any transaction, regardless of its form, that in substance creates a security interest.[43] The Canadian provincial Personal Property Security Acts are similar.[44] Article 9 and the PPSAs provide that an unregistered ('unperfected') security interest is ineffective against the debtor's trustee in bankruptcy.[45] The primary object is to maintain the integrity of the register by giving the secured party an

[37] (2000) 171 ALR 568 at para 40. That was not quite the end of the matter. The contract between Seller and Buyer gave Buyer credit for the price until November 1995. Buyer's liquidator argued that the real purpose of the proceeds subclause was to give seller a security interest in the resale proceeds to secure payment of Buyer's debt. If the real purpose was to create a security interest, that is inconsistent with an intention to create a trust. The majority responded by finding an implied term in the contract to the effect that the debt was discharged the moment the trust arose. The trust arose when Buyer received the sub-buyer's payment. At that point, Seller lost its debt claim and acquired a beneficial interest in the resale proceeds instead. Seller's beneficial interest did not arise until the debt was discharged and so it could not be security for the debt. The terms of the trust were that Seller, as beneficiary, could call on the trust property, but not until after November 1995. In other words, what was originally an express term of the contract (the provision for credit) became incorporated as an express term of the trust when the debt was discharged.

[38] (2000) 171 ALR 68 at para 47.

[39] (2000) 171 ALR 68 at para 47.

[40] (2000) 171 ALR 68 at para 51. In the end Seller lost, but only because it was unable to trace the sub-buyer's payments in Buyer's hands to any particular invoice and batch of steel. In other words, Seller could not prove that the payments were 'proceeds' within the meaning of the proceeds subclause.

[41] (2000) 171 ALR 68 at para 51.

[42] (2000) 171 ALR 68 at para 92.

[43] Revised Art 9, Uniform Commercial Code, s 9.109(a)(1).

[44] See, eg, Personal Property Security Act RSO 1990, c P–10, s 2(a).

[45] See, eg, Personal Property Security Act RSO 1990, s 20(1)(b).

incentive to register its security interest. Kirby J's judgment would move Australian law closer to the Article 9/PPSA position in this respect. However, in other respects it would move Australian law further away.

Consider the following variation of the *Associated Alloys* facts.

> Buyer takes delivery of the steel pursuant to the agreements with Seller. It uses the steel to make steel products which it resells to Third Party. Third Party pays in full on delivery. Buyer defaults in its obligation to pay the price. Seller claims the proceeds from Buyer's sale of the steel products. At all relevant times, Bank holds a fixed and floating charge over the whole of Buyer's undertaking. Bank's security interest is registered under the Corporations Law. Bank appoints a receiver pursuant to the security agreement. The receiver claims the Third Party resale proceeds as part of Bank's floating charge collateral.

Who has priority over the proceeds, Seller or Bank? The Article 9/PPSA position is to say that a purchase money security interest in inventory (like Seller's) or its proceeds has priority over any other security interest in the same collateral (like Bank's) provided: (1) Seller perfects its purchase money security interest before Buyer takes possession of the collateral; and (2) before taking possession, Seller gives Buyer a notice in writing to say that Seller has a purchase money security interest.[46] Seller's perfected purchase money security interest in the inventory (the steel) continues in the end product of Seller's manufacturing process (the steel products) and in the proceeds of the end product (the purchase price Third Party pays Seller).[47] Compare *Associated Alloys*. The majority judgment implies that Seller beats Bank but without the need to register its security interest or notify Bank. A trust is not a registrable charge within the meaning of the Corporations Law and so the statutory priority rules do not apply. The common law rule favours Seller. The resale proceeds are impressed with a

[46] See, eg, Personal Property Security Act RSO 1990, s 33(1). 'Purchase money security interest' means, in part, a security interest taken or reserved in collateral to secure payment of all or part of its price: *Ibid*, s 1(1). 'Proceeds' means 'identifiable or traceable personal property in any form derived directly or indirectly from any dealing with collateral': *Ibid*.

[47] *Ibid*, s 37 ('a perfected security interest in goods that subsequently become part of a product or mass continues in the product or mass if the goods are so manufactured, processed, assembled or commingled that their identity is lost in the product or mass').

The rule stated in the text applies if the proceeds are non-accounts proceeds. In the example under discussion, the proceeds are non-accounts proceeds because Third Party pays in full for the steel products on delivery. In cases involving accounts proceeds (eg, Seller sells the steel products to Third Party on 30, 60 or 90 day terms), the position is more complicated. Under Art 9 itself, the purchase money security interest priority rule in Seller's favour does not extend to accounts proceeds: Revised Art 9, Uniform Commercial Code, s 9.304(a). Some Canadian provinces, for example Saskatchewan, take the same position: Personal Property Security Act SS 1993, c P–6.2, s 34(4). In some other provinces, for example New Brunswick, Seller's special priority does extend to accounts proceeds, but if there is a competition between Seller and a dedicated accounts receivable financier, Seller must notify the financier of its purchase money security interest before Buyer obtains possession of the goods: Personal Property Security Act SNB 1993, c P–7.1, s 34(2). In Ontario, Seller's special priority extends to accounts proceeds and Seller beats a dedicated accounts receivable financier without the need trust from the moment they come into Buyer's hands and so there is no point in time at which Buyer acquires a beneficial interest in them. Since Buyer has no beneficial interest, there is nothing in the proceeds (apart from bare legal title) to which Bank's security interest can attach.[48] On the other hand, Kirby J's minority judgment implies that Bank beats Seller. Seller's interest in the resale proceeds is a registrable charge and so the Corporations Law applies. Assume Seller's charge is unregistered. Bank wins because the Corporations Law says that a registrable charge that is registered has priority over a registrable charge that is not. Now assume Seller's charge is registered. Bank still wins, because the Corporations Law says that as between competing registered charges, priority turns on the order of registration. The majority judgment is closer than Kirby J's to the Article 9/PPSA position. On the other hand, Kirby J's judgment is consistent with the trend in the United Kingdom, where the courts by and large have been hostile to extended reservation of title clauses.[49] It is not possible to assess the relative merits of these competing positions in purely doctrinal terms. Explicit reference to policy considerations—instrumentalism—is unavoidable.

What does the law and economics approach have to say? Baird likens personal property security registration laws to the law governing mineral claims because they make the existence of property rights public in a clear and unambiguous way:[50]

> [a] prospector in the Old West could explore land, confident that no-one else had a claim to it that could be superior to his as long as no markers indicated it belonged to someone else. Similarly, a creditor can lend money to a debtor with confidence because he knows (by looking at the files) that no other existing creditor can claim an interest in the property superior to his own.

A registration system coupled with a first to register priority rule facilitates secured lending at large.[51] A priority rule that favours Seller over Bank inhibits secured lending at large because it reduces the reliability of the register as a source of information to Bank about potentially prior ranking claims and it increases Bank's lending risks. On the other hand, a rule

for prior notice: Personal Property Security Act RSO 1990, c P–10, s 33(1). The policy differences have to do with the relative merits of facilitating trade credit (inventory financing) at the expense of accounts receivables financing (including securitisation arrangements), and *vice versa*: see Jacob S Ziegel and David L Denomme, *The Ontario Personal Property Security Act: Commentary and Analysis* 2nd edn (Toronto, Butterworths, 2000), pp 287–288; Ronald CC Cuming and Roderick J Wood, *British Columbia Personal Property Security Handbook* 4th edn (Toronto, Carswell, 1998), pp 251–252.

[48] The analysis is the same as for simple reservation of title clauses: Michael G Bridge, *The Sale of Goods* (Oxford, Clarendon Press, 1997), p 104.

[49] *Ibid*, p 108–110 and see *Associated Alloys* (2000) 171 ALR 568 at para 67 per Kirby J.

[50] Douglas G Baird, 'Notice Filing and the Problem of Ostensible Ownership' (1983) 12 *Journal of Legal Studies* 53 at p 63.

[51] There is a protracted debate in the law and economics literature about whether the law should encourage secured lending at all. The conventional explanation for secured lending is

favouring Seller facilitates trade credit in particular. Does the benefit of a rule favouring Seller exceed the cost? The most widely accepted view is that a rule favouring Seller is necessary to prevent Bank's situational monopoly over Buyer's financing requirements.[52] The argument goes like this. If the first to register priority rule always applied, Bank would have a monopoly advantage over Buyer's other prospective creditors with respect to future loans. Why? Because Bank's floating charge covers all Buyer's present and after-acquired assets. It leaves no unencumbered assets for Buyer to offer as collateral to a new creditor ('Newcomer'). Newcomer's main choices, aside from refusing to deal with Buyer at all, are to: (1) take a second ranking security interest behind Bank; or (2) grant Buyer an unsecured loan. Either way, Newcomer will want to charge Buyer a higher interest rate to cover its increased risk. But an increase in Newcomer's interest rate is likely to make Newcomer uncompetitive with Bank. Therefore Buyer will choose to deal with Bank. Why would Buyer agree to submit itself to Bank's situational monopoly in the first place? The answer must be that Bank compensates Buyer by offering Buyer a lower interest rate.[53] Therefore, a rule to prevent Bank's situational monopoly will increase the interest rate that Buyer pays Bank. In other words, the right to give later lenders priority over Bank can only be acquired at a price. Presumably Buyer would not agree to pay the price unless there was an offsetting gain, namely that the premium Bank charges Buyer for the first loan is less than the savings in interest rates Buyer can make by offering special priority status to a later financier (Seller). According to Schwartz, this offsetting gain is most likely to be present in

that higher risk debtors prefer to grant security because it enables them to borrow and creditors prefer to take it because it enables them to make loans they would otherwise refuse. Cf Alan Schwartz, 'Security Interests and Bankruptcy Priorities: A Review of Current Theories' (1981) 10 *Journal of Legal Studies* 1, arguing that if creditors are homogeneous, risk neutral and perfectly informed, secured lending is a zero sum game. Under these conditions, the benefit to one secured creditor from taking security (reduced risk) will be exactly offset by the increased cost imposed on an unsecured creditor whose claim to the debtor's asset pool is *pro tanto* diminished. Correspondingly, the benefits to the debtor from granting security (a reduced interest rate) will be exactly offset by the higher interest rates that unsecured creditors will demand by way of compensation for the increased risk they now face. The debtor's total interest bill is thus unaffected by the existence of security. Furthermore, since there are transactions costs associated with the granting of security, the debtor is actually worse off with security than without it. See also Alan Schwartz, 'A Theory of Loan Priorities' (1989) 18 *Journal of Legal Studies* 209 and Lucien Arye Bebchuk and Jesse M Fried, 'The Uneasy Case for the Priority of Secured Claims in Bankruptcy: Further Thoughts and a Reply to the Critics' (1997) 82 *Cornell Law Review* 1279. These are the leading contributions to the debate. Citations to many other contributions appear in the Schwartz and Bebchuk and Fried articles.

[52] Thomas H Jackson and Anthony T Kronman, 'Secured Financing and Priority Among Creditors' (1979) 88 *Yale Law Journal* 1143. There is an alternative argument in support of a rule favouring Seller. This says that since Seller's transaction with Buyer brings new money into Buyer's business, Bank is no worse off and may be better off (each dollar of debt Buyer incurs to Seller is offset by a dollar of new asset value). For criticism of this argument, see Schwartz, 'A Theory of Loan Priorities', *op cit* at p 229.

[53] Jackson and Kronman, *op cit* p 1171.

cases where Seller's finance agreement with Buyer is a purchase money finance agreement (in other words, the purpose of the agreement is to finance the acquisition of a particular asset). There are three reasons:

(1) purchase money loans are often used to continue the project Bank initially approved, and not to finance new, riskier projects. So long as the initial loan is not in default, Bank will want Buyer to continue its existing projects in order to generate revenue to make repayments. Therefore, Bank is more likely to demand a lower premium in exchange for giving Buyer the right to confer special priority status on a purchase money security lender (Seller) than where the loan is for other purposes.
(2) Having other lenders make purchase money loans may reduce total credit costs. Initial lenders commonly extend large sums to enable debtors to pursue major projects. Firms that make such loans often do not want to make numerous small commercial loans that do not require serious credit evaluations.
(3) Some assets are sufficiently specialised that sellers have a comparative advantage over general purpose lenders at maximising the proceeds from resale after repossession, and this should be reflected in lower interest rates.

For these reasons, Bank and Buyer are more likely than not to agree that Buyer should have at least limited authority to grant prior ranking purchase money security interests to third parties. Why not simply leave it to Bank and Buyer to arrive at this result themselves by negotiation—is a statutory priority rule in Seller's favour really necessary? If a bargaining outcome is more likely than not, then it is cheaper to anticipate it by legislation. The Article 9/PPSA purchase money security interest priority rule is a kind of statutory implied term that the parties are free to exclude.[54] The rule saves transactions costs because it avoids the need for parties in the majority of cases to bargain expressly for a rule that favours Seller.

In summary, the situational monopoly argument supports a special priority rule in Seller's favour, subject to any restrictions in the contract between Buyer and Bank. Should Seller have to register its security interest? The answer is 'yes', to prevent prejudice to third parties (other than Bank) who deal with Buyer in reliance on the register—for example, execution creditors and other prospective secured and unsecured creditors. Should Seller have to notify Bank of its purchase money security interest? The purpose of the Article 9/PPSA notification rule is to give Bank fair warning of Seller's purchase money security interest before Bank makes further

[54] See, eg, Personal Property Security Act RSO 1990, c P–10, s 38: 'A secured party may, in the security agreement or otherwise, subordinate the secured party's security interest to any other security interest and such subordination is effective according to its terms'.

advances to Buyer. Seller's notice alerts Bank to the risk that Seller has a prior ranking security interest in Buyer's new inventory and so there may be insufficient collateral to meet Bank's further advances. Bank could discover Seller's purchase money security interest by searching the register before making any further advance. However, this precaution will be a costly one for Bank if there are multiple and irregular advances—for example, where Bank's security interest secures an overdraft or other revolving credit facility in Buyer's favour. The notification rule avoids excessive transactions costs in such cases by making repeated register searches unnecessary.

The explicit difference between Kirby J and the majority in *Associated Alloys* is one of statutory interpretation: what are the limits of judicial creativity? The majority's response was a cautious one and the above analysis supports this position. Kirby J's judgment assumes that, as a matter of policy, Seller's interest should be registrable. This is right as far as it goes. Non-registration of Seller's interest reduces the reliability of the register as a source of information to third parties. However, there are competing considerations at stake. If Seller's interest is registrable then, under the Corporations Law as it presently stands, Bank will beat Seller in a priority dispute whether or not Seller registers. The better rule may be that Seller beats Bank at least if Seller registers. Kirby J's judgment yields the right rule for cases where the dispute is between Seller and Buyer's liquidator and the wrong rule for cases where the dispute is between Seller and Bank.[55] The trade-off is an avoidable one, but only by statutory reform. In cases like this, it is better for the courts not to try and second-guess the legislature.[56]

IV. EQUITABLE REMEDIES IN BANKRUPTCY

Damages are the usual remedy in tort and contract. The equitable remedies of injunction and specific performance are available, but the plaintiff has to show that damages are inadequate.[57] If the defendant is bankrupt, it could be argued that the plaintiff should always be entitled to specific relief because damages are an inadequate remedy when the defendant cannot pay. On the other hand, it could be argued that the plaintiff should never be

[55] But not necessarily if, contrary to the fact situation under discussion, the resale proceeds are accounts proceeds: see n 47, above. In that case, the benefits of facilitating securitisation arrangements and the like may outweigh the benefits of facilitating trade credit. In any event, this, too, is a question that the legislature is better placed than the courts to resolve. Courts are unlikely to have the information necessary to make the assessment.

[56] See: Law Commission, Registration of Security Interests: Company Charges and Property Other Than Land (Consultation Paper No 164, 2002), provisionally recommending reforms to English law along the lines of the Art 9/PPSA model; Australian Law Reform Commission, Personal Property Securities (Report No 64, 1993), making recommendations along similar lines; and Personal Property Security Act 1999 (NZ), which is based substantially on the Saskatchewan and New Brunswick Canadian provincial PPSAs.

[57] RP Meagher, WMC Gummow and JR Lehane, *Equity Doctrines and Remedies* 3rd edn (Sydney, Butterworths, 1992), p 498.

entitled to specific relief if the defendant is bankrupt because specific relief satisfies the plaintiff's claim in full and correspondingly diminishes the entitlements of the defendant's other unsecured creditors. It seems generally to be agreed that the 'always' and 'never' arguments are both untenable.[58] Sometimes is the preferred position. The challenge is to define the sometimes and on this question opinions differ.

A common response is to say that the availability of specific relief against a bankrupt defendant depends or should depend on whether the plaintiff has property rights in the disputed asset. If so, the remedy should be available to protect the plaintiff's entitlement. If not, the plaintiff should be limited to a damages claim which would be admissible to proof in the defendant's bankruptcy. For example, Goode says that in the case of a contract for the sale of land, the buyer becomes the owner in equity upon exchange of contracts and so the buyer can obtain specific performance upon tender of the price even if the seller is bankrupt.[59] By contrast, he says, in the case of a contract for the sale of goods, the sale of goods legislation leaves no room for the vesting of equitable title in the buyer unless the parties themselves have expressly created a trust or equitable interest:[60]

> So a buyer who has paid the price before ownership or possession have been transferred to him cannot expect to salvage his position by obtaining an order for specific performance against his seller, for since he has no proprietary interest in the goods the effect of an order for specific performance would be to remove an asset from the general body of creditors for no corresponding benefit and give the buyer preferential treatment in breach of the *pari passu* rule.

This statement assumes that the plaintiff's proprietary interest is independent of the remedy to enforce it or, in other words, that the 'entitlement' question is separate from the 'remedy' question. In truth, the two questions are interdependent. Property rights in equity are remedy-driven. If the plaintiff is entitled to specific relief, equity anticipates the outcome and treats the plaintiff as having a proprietary interest in advance of the order ('equity deems as done what ought to be done'). Therefore, it is at least as true to say that the remedy justifies the proprietary interest as it is to say that the proprietary interest justifies the remedy.[61] In the land case, for example, one can say that the buyer becomes the owner in equity when

[58] *Ibid*, pp 526–527; I C F Spry, *The Principles of Equitable Remedies* 6th edn (Sydney, LBC Information Services, 2001), pp 68–69; R M Goode, *Principles of Corporate Insolvency Law* 2nd ed (London, Sweet & Maxwell, 1997), pp 124–127.
[59] *Op cit* p 126.
[60] *Ibid*, p 127. For a similar analysis, see Jay L Westbrook, 'A Functional Analysis of Executory Contracts' (1989) 74 *Minnesota Law Review* 227 at pp 257–263.
[61] See Spry, *op cit* at p 664:

> It is often said, where there is a contract for the sale of land, for example, that in equity property in that land passes to the purchaser; and, indeed, undoubtedly, he may often protect his interest by obtaining an injunction or other such equitable remedies. But in

contracts are exchanged and so the buyer is entitled to specific performance. Alternatively, one can say that specific performance becomes available upon exchange of contracts and so the buyer is the owner in equity. In the sale of goods case, a buyer cannot normally claim specific performance. Damages are an adequate remedy so long as there is an available market for the buyer to purchase substitute goods. However, the sale of goods legislation allows for specific performance in exceptional cases,[62] and there can be little doubt that where specific performance is available the buyer does have an equitable proprietary interest in the goods.[63] In short, it begs the question to say that the availability of specific relief in bankruptcy turns, or should turn, on whether the plaintiff has a proprietary interest in the disputed asset.

What does the law and economics approach have to say? According to Jackson, the purpose of the bankruptcy laws is to solve a common pool problem.[64] Outside bankruptcy, as a general rule unsecured creditors are entitled to satisfaction of their claims on a first-come, first-served basis. The first-come, first-served rule gives individual unsecured creditors an incentive to stake their claims as quickly as possible. Being first increases the chances of being paid. On the other hand, a race to judgment between individual creditors is likely to result in the piecemeal dismantlement of the debtor's estate. This will be contrary to the interests of the creditors as a group if the debtor's assets are worth more together than separately. The bankruptcy laws maximize the value of the debtor's estate to the unsecured creditors as a group by substituting a system of collective debt collection for the individual debt collection system that operates outside bankruptcy. The laws conserve the debtor's estate by restraining self-interested behaviour on the part of individual creditors, in much the same way as fisheries laws conserve fish stocks by restraining over-fishing on the part of individual fishers.

Premature bankruptcy is against the interests of the creditors collectively. While the debtor remains solvent, there is always the prospect that it will be able to trade its way out of trouble. Putting a firm into bankruptcy reduces the prospect of all creditors being paid in full because bankruptcy

> truth the statement that the purchaser has acquired equitable property can mean no more in this context than that he has acquired certain equitable rights to the specific performance of the contract … So the vendor is, in those circumstances, required to hold for the purchaser, or exercise in his favour various rights that he may be found to possess, whether the contract in question relates to land or anything else of a tangible or intangible nature, provided, of course, that there is a right to specific performance.

[62] See eg, Sale of Goods Act 1979, s 52(1).

[63] See eg, *Re Wait* [1927] 1 Ch 606 at pp 634–635 *per* Atkin LJ:

> In the view that I have taken of the facts of this case, I have already said that the goods were never so ascertained that specific performance would ever have been ordered of them. This consideration would appear to defeat the supposed equitable assignment.

[64] Thomas H Jackson, *The Logic and Limits of Bankruptcy Law* (Cambridge Massachusetts, Harvard University Press, 1986), ch 1.

proceedings set the sharks circling: suppliers are liable to refuse further credit, the bank will want to call in its overdraft, and so on. To reduce the incidence of premature bankruptcies, creditors' relative entitlements should be the same inside and outside bankruptcy. If Creditor A improves its position relative to Creditor B in Debtor's bankruptcy, Creditor A has an incentive to use the bankruptcy laws opportunistically—to put Debtor into bankruptcy for its own benefit even if bankruptcy is not in the interests of the creditors collectively. Changing creditors' relative entitlements in bankruptcy conflicts with the collectivisation goal of the bankruptcy laws because it encourages self-interested behaviour on the part of individual creditors whereas the goal is to subordinate individual self-interest to the interests of the group. Spry says that the availability of specific relief inside bankruptcy turns on 'whether the inadequacy of damages can be established on some other basis than the defendant's impecuniosity'. In that case, 'it would ordinarily be expected that specific performance would be awarded despite consequent prejudice to other creditors of the defendant'. By contrast, in other cases, 'where the inappropriateness of damages depends only on the defendant's impecuniosity', it may 'be unjust to allow the plaintiff this advantage over other creditors'.[65] In other words, the plaintiff's claim to specific relief in the defendant's bankruptcy should normally be the same as outside bankruptcy. This analysis is broadly consistent with Jackson's theory.[66]

The analysis is the same for other equitable remedies, for example the constructive trust. In Canada and the United States, the main purpose of the constructive trust remedy is to prevent unjust enrichment.[67] Sherwin analyses constructive trusts in bankruptcy from a corrective justice perspective.[68] She argues that different considerations apply depending on whether or not the defendant is bankrupt. If the defendant is not bankrupt, the question is one of unjust enrichment between the plaintiff and the defendant itself. If the defendant is bankrupt, the question is one of unjust enrichment between the plaintiff and the defendant's unsecured creditors: does the inclusion of the disputed asset in the bankrupt defendant's estate unjustly enrich the

[65] *Op cit* pp 68–69.

[66] There is a substantial body of law and economics literature on the relative merits of specific relief and damages outside bankruptcy. For a survey of the literature on specific performance, see Thomas S Ulen, 'Specific Performance' in Peter Newman (ed), *The New Palgrave Dictionary of Economics and the Law* (London, MacMillan Reference Limited, 1998), v III, p 481. On injunctions see especially Guido Calabresi and A Douglas Melamed, 'Property Rules, Liability Rules and Inalienability Rules: One View of the Cathedral' (1972) 85 *Harvard Law Review* 1089 and other literature surveyed in Emily Sherwin, 'Calabresi, Guido' in *The New Palgrave Dictionary of Economics and the Law, Op cit* v I, p 199.

[67] See, eg, *Pettkus v Becker* (1980) 117 DLR (3rd) 257 (SCC).

[68] Emily L Sherwin, 'Constructive Trusts in Bankruptcy' [1989] *University of Illinois Law Review* 297. See Also David M Paccioco, 'The Remedial Constructive Trust: A Principled Basis for Priority Over Creditors' (1989) 68 *Canadian Bar Review* 315, for a similar analysis in a Canadian context.

other creditors at the plaintiff's expense? According to Sherwin, the answer depends on whether the plaintiff voluntarily accepted the risk of the defendant's bankruptcy. If so, there is no principled reason for treating the plaintiff better than other creditors. If not, then on corrective justice grounds the plaintiff is entitled to be preferred.

Sherwin's analysis envisages cases where the outcome of a constructive trust claim may be different depending on whether the defendant is bankrupt. Adapting the facts of *Sherwood v Walker*,[69] she says:[70]

> suppose the debtor and the claimant, a dealer in cattle, agreed to the sale of a pedigreed cow at a low price. Both parties believed the cow was barren. After the claimant delivered the cow, it became evident that the cow was not barren and was worth ten times the price the buyer paid. Outside bankruptcy, assuming a definite mistake by both parties, the seller probably could rescind the sale and obtain restitution of the cow on the ground of a mutual mistake in the basic assumption underlying the contract.
>
> In the buyer's bankruptcy, though, the seller of the cow does not have a strong case for priority over general creditors … . The seller entered into a commercial exchange with the debtor, and a mistaken assumption of facts about the cow may be a hazard of the trade … the risk of such a mistake, in this context, is very close to a voluntary credit risk. The unequal exchange resulting from an honest and reasonable mistake may justify unwinding the transaction when restitution affects only the original parties. But if the mistake was an inherent risk of a commercial transaction, and the restitution claim arises in bankruptcy, the claimant is not fairly distinguishable from other creditors who suffered a loss because they misjudged the debtor's creditworthiness.

Constructive trust orders are sometimes made to deter conflicts of interest and other fiduciary wrongdoings rather than to prevent unjust enrichment.[71] *Attorney-General for Hong Kong v Reid*[72] (employee bribe-taking) and *Boardman v Phipps*[73] (use of fiduciary office for personal gain) are good examples. Sherwin argues that in cases like this, the constructive trust claim should fail if the defendant is bankrupt.[74] The deterrence rational disappears in bankruptcy because the court's order has no impact on the debtor personally and there is no justification on either deterrence or corrective justice grounds why the creditors should bear the loss.[75]

Economic analysis focuses on the costs of reasoning like this. Changing creditors' relative entitlements inside bankruptcy is an incentive for

[69] 66 Mich 568, 33 NW 919 (1887).
[70] *Op cit* at pp 359–360.
[71] *Soulos v Korkontzilas* [1997] 2 SCR 217.
[72] [1994] 1 AC 324.
[73] [1967] 2 AC 46.
[74] *Op cit* at pp 337–339.
[75] See James Edelman, 'Remedial Certainty or Remedial Discretion in Estoppel After *Giumelli*?' (1999) 15 *Journal of Contract Law* 1 at p 54.

individual creditors to use the bankruptcy laws opportunistically. Creditors in competition with the constructive trust claimant will want to put the defendant into bankruptcy so as to extinguish the constructive trust claim. The likely consequence is an increase in the number of bankruptcies and losses to creditors as a group and society at large. Contrary to Sherwin's own summary assessment,[76] it is questionable whether the gains on the corrective justice front are sufficient to offset the losses on the efficiency front.[77]

V. CONCLUSION

Mason rejects instrumentalism and in particular the law and economics approach as a basis for judicial decision-making. He favours the deciding of cases 'within the framework of traditional legal principles' and without reference to policy considerations. This paper has attempted to refute Mason's argument. Non-instrumentalist reasoning is a recipe for judicial error. The courts cannot hope to get the law consistently right unless they address the consequences of their decisions. All legal interventions have costs as well as benefits. How can a court possibly know whether a decision it proposes to make is a good one unless it identifies the potential costs? On the other hand, as soon as the court sets foot down this path, it is embarked on economic analysis, though perhaps at a very basic level.

Observation suggests that courts, or superior courts at any rate, do not routinely behave in the way Mason suggests they should. Courts do tend to reason instrumentally, even if their policy analysis is rudimentary and not very explicit. The cases discussed above bear out this observation. In *O'Brien's* case,[78] the court adopted an overtly instrumentalist approach. In *Pilmer's* case,[79] the court's policy analysis was less explicit but it is clear that concerns about the potentially chilling effect on professional relationships of a decision in the plaintiff's favour was an important factor in the majority's decision. In the *Associated Alloys* case,[80] the majority noted the potentially adverse policy consequences of its decision, but it said the problem called for a legislative solution. The court could not fix it. This turns out to have been a wise response because there are competing values at stake in the registration of reservation of title clauses. These go to the design of the registration statute and the court will not necessarily have sufficient

[76] *Op cit* p 366.
[77] From this perspective, a better approach to cases like *Attorney-General for Hong Kong v Reid* is to deny constructive trust relief altogether and limit the claimant to an account of profits: *Lister v Stubbs* (1890) 45 Ch D 1. See, eg, Graham Virgo, *The Principles of the Law of Restitution* (Oxford, Clarendon Press, 1999), pp 540–544; Edelman, *Op cit* pp 54–55.
[78] [1994] 1 AC 180.
[79] (2001) 75 ALJR 1067.
[80] (2000) 171 ALR 568.

information to identify or resolve them. Equitable proprietary remedies have important implications for bankruptcy policy. The courts have long been aware of this. To address the potential bankruptcy implications of a proprietary remedy is to engage in instrumental analysis, whether the court in the end grants the remedy anyway (*Reid's* case[81]) or not (*Lister v Stubbs*[82]).

It was suggested in Part 1, above, that if there is a gap between commercial law as expounded and professed by judges and academics and the commercial expectations of the business community, it is most likely the product of a muddling through strategy. To earn the respect of the business community, the courts must be instrumentalist. The cases studied in this paper are a selective sample, but they tend to suggest that at the higher levels, the courts are not just muddling through. Superior courts do reason instrumentally, though some may do it more overtly than others. By and large, it is hard to see why the business community should be dissatisfied with the courts' performance in the commercial law arena. There is also, of course, the question of commercial law statutes and the business community's assessment of the *legislature's* performance. Is there a mismatch between the business community's expectations and the legislature's efforts in areas such as company law, cost of credit disclosure, product liability, unfair contract terms, occupational regulation, and so on? There may be quite a different story to tell in this regard, but the story needs a paper of its own.[83]

[81] [1994] 1 AC 324.
[82] (1890) 45 Ch D 1.
[83] For discussion of some of the issues in the consumer protection context, see A J Duggan, 'Some Reflections on Consumer Protection and the Law Reform Process' (1991) 17 *Monash University Law Review* 252.

Commentary on 'Commercial Law and the Limits of the Black Letter Approach'

DAVID GOLD

WHEN THEIR LORDSHIPS' House had to consider whether the London Borough of Hammersmith could enter into weird and wonderful derivative instruments as a power incidental to their other functions,[1] their Lordships suddenly saw the light when Mr Peter Scott QC, Counsel for Hammersmith and Fulham, used in argument the example of the under treasurer at Lincoln's Inn being able to deal with Benchers' money by acquiring derivative instruments as a power incidental to his other functions. It then became an easy decision for their Lordships to determine that the entry into derivative instruments was ultra vires of local authorities' powers. Here the House of Lords were clearly taking into account economic considerations (some would say their own!) when interpreting the relevant statute.

Is it right that judges should give consideration to economic and commercial issues or should they instead, as Professor Duggan asks, simply decide cases 'within the framework of traditional legal principles?' Professor Duggan points out that there is a close inter-connection between law and economics. One cannot ignore economic considerations where there is a choice to be made. The courts have sufficient ability to take economic issues into account without needing expert assistance. And the courts are doing it already, as Professor Duggan points out when he refers to the competing interests of Mrs O'Brien and the bank in *Barclays Bank plc v O'Brien*.[2]

In his paper, Professor Duggan considers whether there is a misalignment between commercial law, as expounded and expressed by judges and academics, and the commercial expectations of the business community.

[1] *Hazell v Hammersmith and Fulham LBC* [1992] 2 AC 1.
[2] [1994] 1 AC 180.

He suggests that if there is such a gap, it is most likely the product of a 'muddling through strategy'. In other words, rather than there being a clear policy universally adopted by the courts, consideration of economic issues is very much 'hit and miss', creating uncertainty amongst those affected by the judgments being given. Professor Duggan's answer is that the courts must address the economic consequences of their decisions.

I have to say that I have doubts as to the wisdom of this view. In developing his theme Professor Duggan considers three commercial law topics:

1. Fiduciary obligations in commercial relationships;
2. Reservation of title clauses in sales contract—Romalpa clauses; and
3. Equitable remedies in bankruptcy.

In looking at fiduciary obligations in commercial relationships, Professor Duggan discusses the Australian High Court case, *Pilmer v The Duke Group (in liquidation)*,[3] where consideration was given to whether an auditor owed a fiduciary duty to the company, the argument being that if he did, there could be a breach of fiduciary duty if the auditor had a continuing relationship with the company. Whilst this was rejected by the majority, Kirby J certainly thought that there could be a conflict in such a case even if, for example, the continuing relationship was simply retaining the job of auditor. The majority envisaged that conflicts could arise if the continuing relationship was a more significant one, such as the provision of consulting services.

Absent a statutory rule, is it for the court to determine that an auditor is a fiduciary and therefore cannot act in certain cases? Should the court take into consideration the financial consequences of such a potential conflict of interest? Is the risk of a conflict a better risk—a cheaper option—than the cost of having to change auditor each year?

Professor Duggan then considered the Australian case of *Associated Alloys*[4] and the existence of a reservation of title clause, which potentially created a conflict between a company's bankers and its unsecured creditors. The issue considered was whether a trust interest in specific property should be extinguished on a company's bankruptcy if the interest had not been registered. In other words, should a document purporting to create a trust be regarded instead as a charge and only be valid if registered?

Is it for the court in determining the legal position of the instrument to have regard to the competing positions of different types of creditor? Is the better course that the legislature should determine who takes priority?

A similar issue arises in the final example used by Professor Duggan, which concerned equitable remedies in bankruptcy. Should the court grant

[3] (2001) 75 ALJR 1067.
[4] *Associated Alloys Pty Limited v CAN 001 452 106 Pty Limited (in liquidation)* (2000) 171 ALR 568.

specific performance of a contract when the consequence of so doing is to deprive unsecured creditors of certain assets? Because of the bankruptcy, no damages award could be enforced and the claimant would rank as an unsecured creditor unless specific performance was ordered.

If the rule was that such a constructive trust claim automatically ended on a company's bankruptcy, there would be an incentive for unsecured creditors to put companies into bankruptcy so as to extinguish the constructive trust claim. That could lead to an increase in bankruptcies. Is that good for society? Should this be a decision for the judiciary or legislature? If the legislature fails to come up with an answer why should the courts, in a sensible and structured way not provide an answer? As long as there is certainty and no muddling through, the business community should be content.

It seems to me that the law and economics approach inevitably requires a trade off between the economic interests of at least two competing interests and perhaps issues of public policy too. Of course, in every action a judge has to decide between the competing interests of the parties. He does so by considering the facts and the law. Why should he not also bring into play issues of economics, which may involve public interest considerations? Is the judge the right person to decide? What is his particular qualification for this task? At least the legislature is accountable. Theoretically, a decision will have been made after proper debate, analysis and consultation. Finally, practitioners and businessmen want certainty. My client does not want me to say that I cannot predict the result because it may depend upon a particular judge's statement on economic issues.

I fear that if judges are required to take economic issues into account when determining cases, however one might seek to avoid it, 'muddling through' is still likely to prevail.

22

The Legal Academy's Contribution to the Development of Commercial Law: An Anglo-Canadian Perspective

JACOB ZIEGEL*

I. INTRODUCTION

THERE ARE LIKELY to be different reactions to the claim implicit in the title of this chapter that institutional scholars in England and Canada have an important role to play in the development of modern commercial law. At one end of the spectrum, non-commercial law academics may wonder what it is about commercial law that distinguishes the work of commercial law scholars from the work done by academics in other branches of private law. I do not find the question very helpful. There may indeed be considerable overlap between the approaches and methods used by scholars in the different branches of private law but this is a long way from saying that they should all be lumped together or that the issues and challenges facing the various branches are the same. So I put this objection aside and hope that the remainder of this paper may sufficiently answer the question.

* This chapter is only exploratory and lays no claim to exhaustiveness. I use 'legal academy' in the title of the chapter as a shorthand expression for legal scholars working in an academic setting. I use 'commercial law' in a very elastic sense as concerned with transactions for gain between merchants, nationally and internationally, banking and financial transactions, commercial insolvency law, and including, at least for illustrative purposes, company law and regulatory aspects of trade and commercial transactions. Similarly, in referring to commercial law scholars, I include those colleagues, particularly restitution scholars, whose work has a direct bearing on important commercial law problems even though their primary interests may reach beyond commercial law. Grateful appreciation to Vaughan Black and Kevin Davis for reading a draft version of this essay and making helpful suggestions, and to Cory Binderup, a Dalhousie law student, and to Nersi Makki, a University of Toronto law student, for research assistance.

The reaction at the other end of the spectrum is distinctly more negative and merits closer examination. It actually has two strains. The first echoes a sentiment long held in England by the bar and the judiciary that legal academics represented the junior branch of the profession and were not expected to contribute much to the development of the law.[1] As I will explain shortly, I believe this anti-academic bias is now on its last legs and hopefully will soon find its complete quietus if it has not already. Certainly, legal academics have been accepted in Canada for more than twenty-five years as an integral part of the legal community and an equal partner with the legal profession. This recognition of the role of the legal academy has been true in the United States for a much longer period, so much so that senior professors at leading US law schools easily rank in prestige with members of the federal judiciary.

The second negative strain may appear a little more robust but also turns out on closer inspection to be devoid of merit. The argument that I ascribe to this school of thought is that commercial law is an intensely practical and untheoretical subject, not suitable for academic speculation or rumination, and that little can be expected from legal academics that will genuinely advance understanding or progress in this branch of the common law. This reasoning may be buttressed by the claim that historically commercial law in common law jurisdictions has developed without the help of legal academics and that there is no reason why commercial law should be less self-sufficient in the future. So, it may be urged, by way of example, the foundational principles of contract, sales, agency, banking, insurance, admiralty, and personal property law all evolved without the aid of law faculties and their incumbents, and are none the worse because of it.

The suggestion that commercial law was at any time, and particularly in its most formative periods, free of external influences seems to me highly contestable.[2] So is the implication that the common law doctrines and rules were always rational and efficient and therefore not in need of change and improvement. In any event, the reasoning is flawed because it proceeds

[1] Cf PS Atiyah and Robert S Summers, *Form and Substance in Anglo-American Law* (Oxford, Clarendon Press, reprint 1997), ch 14, p 407 (claiming that there is in England a common legal culture which does not really include law schools at all and that they are viewed as outsiders, both geographically and intellectually.) For a striking and much earlier illustration of this point, compare the law lords' reaction to Jim Gower's complaint in his inaugural lecture at the LSE in 1950 that the judges patronised law teachers in public addresses to them. See 'Jim Gower—An Appreciation' (1998) 61 *Modern Law Review* 127, 131. Happily, as I note later in this chapter, relations between Her Majesty's judges and English academics have vastly improved since then.

[2] To cite some random examples of 19th century external influences, the utilitarians exercised great influence on some English judges, and the writings of Anson, Dicey and Pollock were of great importance respectively in clarifying and systematising contract, conflict of laws, and partnership law principles. (Pollock was responsible for the drafting of the Partnership Act, 1890). In the US, even greater influence was exercised by the heroic labours of Story and Kent in the first half of the 19th century, and by such luminaries as OW Holmes and Samuel Williston in the second half.

from the negative premise that *because* until comparatively recently there were few Commonwealth law schools and few commercial law scholars working in an academic environment, this proves the dispensability of the academic contribution. The proponents of this line of reasoning also appear to assume that the leisurely pace at which, for the most part, commercial law developed in earlier centuries is still adequate for the post-industrial cyberspace and globally driven society that will almost certainly characterise the twenty-first century.

II. THE GREAT TRANSFORMATION

There is a wealth of evidence that contradicts these assumptions. Consider, among others, the following developments over the past twenty-five to fifty years. First, commercial law has grown greatly in importance given the rapid increase in domestic and global trade, the establishment of the European Union, the North American Free Trade Agreement (NAFTA) and other regional trade pacts, the work of the WTO, and the impact of the IMF and the World Bank. Commercial law has become much more specialized and many new branches of commercial law, have evolved that were unknown or barely visible twenty-five years ago.[3]

Similarly, economic, technological, trade and institutional developments have sparked the need for new laws, and new approaches and solutions to old problems to address the emerging challenges, both domestically and internationally. Although they enjoy no monopoly, legal scholars are well situated to answer the call because they have the time, the inclination and the specialised knowledge that may not be available to or shared by other members of the legal community.

In both Canada and the UK there has been a rapid growth in the number of law schools and with it the need for many more law teachers at all levels. In Canada, there are now 21 law schools,[4] though most of the increase occurred in the 1970s. In the United Kingdom, there are about 76 law degree-granting institutions,[5] or about 180 per cent more on a population basis than in Canada. The proliferation of legal academies reflects in part the growth of the legal profession and the need for more law graduates, and in part the attractions of law as an academic discipline.

[3] According to my quick count, the Society of Legal Scholars (SLS) *Directory of Members, 2002*, pp 145ff, lists an astonishing 386 special interests held by its members.
[4] Information supplied by Lily Leung, reference librarian, University of Toronto law library.
[5] This number was supplied to me by Peter Birks in an email of 16 January 2003. However, in making the numerical comparison between Canada and the UK it is important to bear in mind that most Canadian law faculties have professional status and their law degrees are recognised by the provincial law societies as qualifying their graduates to practise law subject to satisfying articling requirements and attending bar admission or similar practical training courses.

Law teachers too find much stimulation and many challenges in the new environment.[6] They are many more opportunities for specialization than there were before as well as new pressures and incentives. Law teachers also appreciate that they must demonstrate legal scholarship if they wish to progress through the ranks and be appointed to leading law schools. Commercial law enjoys a very prestigious position at British law schools and therefore attracts a large following of talented teachers and scholars.[7] The Canadian position is more complex because of Canada's federal structure, the importance of constitutional law, and the overarching impact of the Canadian Charter of Rights and Freedoms since 1982. Nevertheless, there can be little doubt that the opportunities for commercial law scholars are as great in Canada as they are in the United Kingdom. Rather, the concern in Canada is that able scholars will be seduced by the much higher salaries offered by law firms compared with the relatively rigid salary structure obtaining at many if not most Canadian law schools.[8]

In my view, the single most important encouragement to scholarship in all branches of the law has been the growing acceptance of legal academia as an equal partner in the legal community. Canada is in advance of the UK in this area, but in both jurisdictions there has been a complete sea change compared with the situation fifty years ago. In both countries, legal academics play leading roles in law commissions and law reform. In Canada, many law professors have been appointed to the federal and provincial bench at all levels of the judicial hierarchy, including notably the Supreme Court of Canada.[9]

[6] Not all Canadian scholars share this enthusiasm. In 1983, The Consultative Group on Research and Education in Law, chaired by Dean Harry Arthurs of the Osgoode Hall Law School, presented a report, *Law and Learning: report to the SSHRC*, that was critical of the lack of intellectualism in many of the contemporary law school offerings. The Report recommended that law students be divided into two streams, one predominantly interested in the theoretical foundations of modern law and the other focussing on preparing students for the practice of law. The recommendations received a cool reception and were not, to the best of my knowledge, adopted at any Canadian law school.

[7] Again basing myself on a quick count, the SLS Directory lists 112 members with particular interests in commercial law, about the same number as for contract law and somewhat more than for company law. The Directory also lists 20 commercial law specialties, not including company law and its offshoots.

[8] Canadian law schools not only have to contend with much higher Canadian law firm salaries but, increasingly, with the still higher salaries and better prospects offered by large US law firms. In the case of their best young scholars, Canadian law schools also face competition from US law schools whose salaries on average are at least fifty per cent higher than those available in Canada. See HA Arthurs, 'Poor Canadian Legal Education: So Near to Wall Street, so Far from God' (2000) 38 *Osgoode Hall Law Journal* 381. To meet these challenges the University of Toronto law school decided in 2001 to implement a five year programme of substantial annual increases in tuition fees, the intention being to earmark a substantial part of the new income to improve faculty salaries. The precedent has since been followed by at least two other law schools (the Osgoode Hall Law School and the University of Western Ontario law school) and may well be followed by other Canadian law schools facing similar competitive pressures unless the tuition hikes are vetoed by provincial governments.

[9] For a now somewhat dated study see PH Rusell and JS Ziegel, 'Federal Judicial Appointments: An Appraisal of the First Mulroney Government's Appointments and the

Academic articles and studies are regularly cited in Canadian judgments at all levels and particularly in judgments of the Supreme Court of Canada.[10] A similar trend has emerged in the UK over the past decade or more though at a significantly slower pace. Also, the frequency of academic citations in English judgments varies much more by subject matter and level of court considering the issue.[11] In Canada, the status of academic writing is greatly enhanced by the fact that possession of a law degree is a prerequisite in all provinces for call to the bar and admission to practice as a solicitor, thereby creating a captive market of generations of (hopefully) appreciative law graduates. Just as important is the introduction of law clerks to assist the judges in the preparation of judgments and review of applications for leave to appeal. The Supreme Court of Canada led the way starting in 1968[12] and the practice has since spread to the lower courts at both the federal and provincial levels. The recent introduction of legal assistants in the House of Lords is also a belated and welcome recognition in the UK that more extensive research into the ramifications of legal problems can only enrich the judgments, and must surely also lead to still greater appreciation of the value of academic scholarship.

Finally, it should be noted that in both Canada and the UK commercial law scholars often represent their countries at international meetings and in the drafting of model laws and international conventions by bodies such as UNIDROIT, UNCITRAL, the Hague Conference on Private International

new Judicial Advisory Committees' (1991) 41 *University of Toronto of Toronto Law Journal* 4, table 6 (showing that 5 full time law professors were appointed to the bench by the Mulroney administration in the period 1984–88. The number of such appointments increased substantially in subsequent years although practising lawyers still constitute the main source of appointments by a wide margin. In the late 1990s, three out of the nine members of the Supreme Court of Canada were former law professors, viz, Iacobucci, LaForest and McLachlin JJ. Justice McLachlin subsequently became Chief Justice of Canada.).

[10] For the role of academic citations in the Supreme Court, see P McCormick, 'Do Judges Read Books Too? Academic Citations By the Lamer Court, 1991–6' (1998) 9 *Supreme Court Law Review* 463 and V Black & N Richter, 'Did She Mention My Name? Citation of Academic Authority By The Supreme Court Of Canada, 1985–1990' (1993) 16 *Dalhousie Law Journal* 377.

[11] Perhaps the most remarkable change in frequency of academic citations has occurred in restitution cases where Prof Andrew Burrows tells us that 'No practitioner worth his or her salt would think of coming before the Court of Appeal or House of Lords on an issue in the law of restitution without being acquainted with the views of the academics, most notably Gareth Jones … and Peter Birks.' A Burrows, 'The English Law of Restitution: A Ten-Year Review', p 1, paper delivered at the University of Western Ontario, 25 January 2003, at a Symposium on *Understanding Unjust Enrichment* and to be published by Hart Publishing (January 2004). For an excellent and more nuanced evaluation of the influence of juristic writing on contemporary English judgments see Neil Duxbury, *Jurists and Judges. An Essay on Influence* (Hart Publishing, 2001).

[12] I am indebted to my former colleague Justice Robert Sharpe of the Ontario Court of Appeal for this information. The Supreme Court of Canada currently hires 27 law clerks annually, three for each of its nine members. See further M McInnes, J Bolton and N Derzko, 'Clerking at the Supreme Court of Canada' (1994) 33 *Alberta Law Review* 58 and L Sossin, 'The Sounds of Silence: Law Clerks, Policy Making and the Supreme Court of Canada' (1996) 30 *University of British Columbia Law Review* 279.

Law and, of course, at the EU level, by the various organs of the EU and the European Parliament. This not only brings the scholars into contact with their counterparts from other jurisdictions but also gives them precious opportunities to learn about other legal systems and to act as intermediaries in mediating differences between common law and civil law concepts in the drafting of international conventions and model laws.

III. THE MANY ROLES OF COMMERCIAL LAW SCHOLARS

It is of course encouraging for commercial law scholars to know that their contributions are valued. It is just as important, I believe, to have a good understanding of the varying roles these academics may be called upon to play and how the roles may be changing. Traditionally, the most common form of scholarly contributions in England and Canada has been in the form of comments, articles, monographs and textbooks, mostly of a doctrinal character. The current Canadian judicial practice is to cite all four sources very generously, some might argue too generously. However, there is a good reason for the practice. Canada is a huge country and judges are scattered across great distances. They do not have the good fortune of their English cousins of being largely concentrated in one metropolitan centre where they can readily consult one another and regularly have the benefit of hearing arguments by leading counsel. In Canada, there is also a great proliferation of reported cases since almost every written decision is reported somewhere. As a result, judges are often swamped with case citations. Given this jurisprudential cornucopia, it is efficient for the judges to look for short cuts. A comment, article or textbook by a respected academic (or practitioner of course) serves this office admirably.

Nevertheless, it must be conceded that not everything written by academics—even by established scholars—has a long shelf life. Atiyah and Summers complained in 1987[13] that too many textbooks published in England were intended for students and were little more than summaries of black letter law. They lamented the fact that not enough of what was being published expanded the frontiers of knowledge or provided new insights into existing issues. I suspect the complaint was already stale when their book was first published. In any event, contemporary leading UK academics agree that scholarly legal output in England in the 1990s was very impressive in quantity, quality and diversity.[14] No doubt one could make the same observation about Canadian scholarly output, but our problems

[13] Above n 1, at 394.
[14] See P Birks in Peter Birks (ed), *Pressing Problems in the Law, v 2, What Are Law Schools For?* pp vi–viii (Oxford, OUP, 1966) and Wm Twining, *Blackstone's Tower: The English Law School.* Ch 6 (London Stevens & Sons 1994).

lie elsewhere. We do not have nearly enough textbooks in the basic branches of law and, despite the vast flood of decisional and enacted legislation at both federal and provincial levels, much of it attracts little attention even in the form of short comments.[15] Academic colleagues are often too busy working on particular projects and in any event prefer to concentrate their energies on major articles or monographs. For the most part Canadian practitioners do not fill the gaps either because writing for publication does not generate billable hours.

IV. COMMERCIAL LAW SCHOLARS AS CRITICS AND LAW REFORMERS

I want, however, to return to the Atiyah and Summers' lament because it bears on a related topic. Very few branches of commercial law are static but are constantly in need of reinventing themselves. Certainly this is true of contract law where almost every basic doctrine has come under close scrutiny over the past twenty years, including the concept of freedom of contract itself.[16] The same is true of company law where, quite recently, the DTI's Advisory Committee on Company Law recognised the need for a basic overhaul of company law concepts.[17] Sale of goods law has received intermittent attention from the Law Commission for twenty-five years or more. However, in the view of many, the patchwork quilt of amendments has gone too far and the time is ripe for a complete revision of the Sale of Goods Act. Still more significant, Professor Goode has argued the need for a British Commercial Code[18] comparable in scope to the Uniform Commercial Code, which has been such an outstanding success in the US. We are told that the DTI has at least expressed interest in exploring the feasibility of the project. We may be sure that if the project receives government blessing commercial law scholars (almost certainly likely to be led by Roy Goode) will play an active role. In Canada, regrettably, law reform, which seemed to have such a promising future in the 1970s and 1980s, has lost its shine and has been reduced to a shadow of its former self.[19]

[15] Alas, I speak with many years of experience as editor in chief of the Canadian Business Law Journal (CBLJ).

[16] See eg, PS Atiyah, *The Rise and Fall of Freedom of Contract* (Oxford, Clarendon Press, 1979); Roger Brownsword, *Contract Law: Themes for the Twenty-First Century* (London, Butterworths 2000); and E McKendrick, 'English Contract Law: A Rich Past, an Uncertain Future' (1997) 50 *Current Legal Problems* 25.

[17] Report of Company Law Steering Committee, *Modern Company Law for a Competitive Economy: Final Report, URN 01/942 and URN 01/943.*

[18] RM Goode, 'Insularity or Leadership? The Role of the United Kingdom in the Harmonization of Commercial Law' (2000) 50 *International and Comparative Law Quarterly* 751, at 761–63.

[19] In the 1970s and 1980s the federal government and many of the provinces had a law reform agency of some description. This is no longer true. The federal Law Reform Commission was

V. LESSONS OF CHATTEL SECURITY LAW REFORM

Of course, not every desirable change in commercial law requires legislative intervention. In fact, since many of the doctrines and rules are of judicial origin they can be changed by judicial fiat. But we know that respect for stability and the doctrine of *stare decisis* have largely precluded the courts from changing direction even when the case for it was very strong. There have been important exceptions to the rule, even in England, and scholarly writings have played a major role in bringing about such changes, especially in the law of restitution.[20] In Canada, the Supreme Court of Canada has over the past twenty-five years shown a greater willingness to overrule even very long-standing precedents[21]—and has sometimes even overlooked them—and it is safe to assume that this trend will continue.

Canadian courts have been encouraged to play this activist role for a number of reasons. One is that, as final arbiters of the meaning of the entrenched Canadian constitution and especially, since 1982, of the Canadian Charter of Rights and Freedoms, they have become very comfortable with the exercise of normative powers. A second reason is the wide dispersal of legislative powers among ten provinces, three territorial governments, and the federal government. The third, and probably most influential, factor is general lack of interest in private law reform at both levels of government because it is not perceived to be a vote winner. This disinterest also accounts for the disbandment of many law reform agencies in Canada or their 'privatization', and a reduction in funding for the few law reform agencies that remain.

Still there is a limit to what even an activist judiciary can achieve. It is not realistic, nor in many cases desirable, for courts to rewrite basic common

abolished by the Mulroney government in 1993 but was revived in 1996 in a different form and with revised terms of reference under the name of Law Commission of Canada. Of the surviving provincial law reform agencies, the Alberta Institute of Research and Law Reform is the most successful and best funded. The British Columbia and Ontario law reform commissions were disbanded in the 1990s and have not been revived. However, the work of the BC Commission has been continued by the Law Reform Institute of British Columbia, a non-profit organisation sponsored by the BC Law Society and other well wishers, and continues to do excellent work on a shoe string budget. For a comprehensive description of the 2001 status of law reform agencies in Canada, see Nova Scotia Law Reform Commission, Report, *A Continuing Need For Law Reform: The Case for the Law Reform Commission of Nova Scotia* (2001). (I am much indebted to Arthur Close, QC, director of the BC Law Reform Institute for assistance in compiling the above information.)

[20] See above n 11.

[21] A seminal example is *Morguard Investments Ltd v De Savoye*, [1990] 3 SCR 1077, a case in which, as widely interpreted, the Supreme Court rewrote the rules for the recognition and enforcement of foreign judgments and found an implicit full faith and credit clause in the Canadian Constitution. Other examples involve the Supreme Court's recognition that unjust enrichment constitutes a separate branch of the common law, abolition of the rule that relief is not available for payments made under a mistake of law, and adoption of the remedial constructive trust in unjust enrichment cases.

law doctrines, and especially not when they have been woven into the fabric of supportive statutory law.[22] If serious weaknesses are perceived in a doctrinal or statutory structure, ultimately they can only be redressed by legislation.

The reform of chattel security law in common law jurisdictions is a particularly telling illustration of this proposition. It is also a story worth retelling because of its transnational character and the interaction between scholars, practising lawyers, and legislatures.

Every student of the field is acutely aware of the great complexity of chattel security law at common law. There are, or were, a multiplicity of security and quasi-security devices, each governed by its own corpus of rules and doctrines and frequently (especially in North America) superimposed statutory requirements. It was the quintessential example of a branch of commercial law ripe for simplification and modernisation. When Allison Dunham and Grant Gilmore, then both young US commercial law scholars, came to review the American position in the late 1940s in their capacities as joint reporters for what became Article 9 of the Uniform Commercial Code, they independently arrived at a simple but dramatic solution. This was that the complexities of the existing law could be resolved by abolishing the congeries of common law, equitable and statutory security devices and replacing them with the generic concept of a security agreement creating a security interest that would be governed by a common set of rules except where functional distinctions between the security devices and the nature of the collateral justified separate rules. In this way Article 9 of the Code was born and became an instant success, not only because of its intrinsic merits but just as importantly because of the overwhelming support it received from the commercial bar and the banking and financial communities.

A. Canadian Developments

Article 9 was imported into Canada in the early 1960s through the work of the Catzman Committee in Ontario.[23] Again, it was close cooperation

[22] This is not to suggest that it has not happened. For some recent examples where Canadian courts have rewritten basic common law rules, see *Bank of America Canada v Mutual Trust Co* (2002) 211 DLR (4th) 385 (SCC) (allowing recovery of compound interest for breach of contract despite a long line of statutorily based Ontario cases to the contrary), *McIntyre v A G Ontario* (2002) 61 OR (3d) 257 (OCA) (reinterpreting Ontario legislation prohibiting contingency fee arrangements between solicitor and client), and *Taylor v Scurry-Rainbow Oil (Sask) Ltd* (2001) 203 DLR (4th) 38 (Sask CA) (rejection of perpetuities rule in oil and gas leases), discussed by GR Hall in 'Was *Final Note* Not the Final Word? *Scurry-Rainbow* and the Continuing Quest for Balance in the Reform of Common Law Rules' (2002) 37 *Canadian Business Law Journal* 229.

[23] See JS Ziegel, 'The Draft Ontario Personal Property Security Act' (1966) 44 *Canadian Bar Review* 104.

between academics, the commercial bar and industry that ensured the success of the transformation of Canadian chattel security law.[24] Today, all the common law provinces and territories have adopted an Article 9 type Personal Property Security Act (PPSA).[25] Even Quebec, with its French based Civil Code, has adopted a modified version of Article 9 in the form of an *hypothèque* on movable property.[26]

B. English Developments

The Article 9 story becomes particularly arresting when we consider the history of the thirty-year effort to reform chattel security law in England. I believe Roy Goode was the first English scholar to take a serious interest in Article 9.[27] He was able to persuade the Crowther Committee, of which he was a member, to recommend the adoption of a similar functionally based law in England.[28] Regrettably, the early reactions from industry and the legal profession ranged from indifference and scepticism to open hostility.

Nothing much appears to have happened between 1971 and 1986, at which time Prof Aubrey Diamond was asked by the Minister for Consumer and Corporate Affairs to review the current legal position in the light of the Crowther Report and to report whether it was now appropriate to take some action. Aubrey Diamond reported in 1989.[29] He supported the Crowther recommendations and also recommended some interim changes to the charges register under the Companies Act. Nevertheless, his report too yielded no tangible results.

The most recent development in England is the publication by the English Law Commission in the summer of 2002 of a very comprehensive and detailed consultation paper[30] in response to a ministerial reference from the DTI. The reference asked the Law Commission, among other questions, to consider the case for a notice filing system for the registration of company charges and quasi-securities as well as reform of the law governing the granting of security and quasi-security interests by unincorporated businesses and individuals. The contents of the Consultation Paper

[24] See JS Ziegel and DL Denomme, *The Ontario Personal Property Security Act Commentary and Analysis,* 2nd edn, Introduction to the First edition, pp lxv *et seq*.

[25] *Ibid*, Introduction to the 2nd edn, lx.

[26] For the particulars see A Grenon (1996) 26 *Canadian Business Law Journal* 391.

[27] Roy Goode and this author collaborated in the 1960s in the writing of *Hire-Purchase and Conditional Sale. A Comparative Survey of Commonwealth and American Law* (Brit Inst of Int & Comp Law 1965). Art 9 and the Canadian developments figured prominently in our study especially in Ch 14 and onwards.

[28] See *Report of the Committee on Consumer Credit*, Part V (1971).

[29] AL Diamond, *A Review of Security Interests in Personal Property* (DTI, 1989).

[30] Law Commission, *Registration of Security Interests: Company Charges and Property Other Than Land*, Consultation Paper No 164 (2002).

were hardly surprising. The paper supported the establishment of a notice filing system for the registration of company charges and equally endorsed the adoption of an Article 9 type regime to regulate security and quasi-security interests given by non-corporate debtors. It is an open secret that academic input played a major role in the preparation of the Law Commission's document.

As an overseas observer, it would be presumptuous for me to predict whether the Law Commission's report will lead to more tangible results than the earlier reports. I have been told however, that the prospects are substantially more favourable because of City of London solicitors' dissatisfaction with the operation of the Companies' register and continuing uncertainties over the distinction between fixed and floating charges over inventory and assignment of accounts and uncertainty over the status of quasi-securities.

Be that as it may, the long delay in the reform of English chattel security law raises important questions about the impact of academic writings and the greater influence of the finance industry and practising lawyers in determining whether, and when, basic changes will be implemented. We may also ask why Article 9 secured such an easy passage in the US in the 1960s compared to the long travails to which its transatlantic counterpart is being exposed in England. These questions will come as no surprise to political scientists, sociologists and economists since they have long accepted the importance of studying group interests in determining the amalgam of factors that lead to the enactment of important new legislation. Perhaps the moral is that academics must be prepared, if the stakes are high enough, to leave their cloistered studies for the legislative halls and to agitate more vigorously for needed reforms.

VI. CHALLENGES OF INSOLVENCY LAW REFORM

I suspect future historians will find it ironic that England is seriously considering adoption of an Article 9 type regime at the very time that there is renewed debate among scholars about the efficiency of secured credit and the priority that should be given secured creditors on a debtor's insolvency. The efficiency question was first raised by Prof Alan Schwartz, a leading US member of the law and economics school, in a much-cited article in the Journal of Legal Studies in 1981.[31] He argued that, viewed abstractly, secured credit was a zero sum game since fully informed unsecured creditors, knowing or anticipating the giving of security, would raise the costs of their own claims, thus cancelling out any benefits the debtor might

[31] Alan Schwartz, 'Security Interests and Bankruptcy Priorities: A Review of Current Theories' (1981) 10 *Journal of Legal Studies* 1.

derive from the reduced cost of secured credit. He also suggested that to the extent some classes of unsecured creditors, eg, involuntary creditors, were unable to adjust their claims this might make a case for compensation by giving them higher priority over the secured creditor's claim. Prof Schwartz's challenging thesis provoked many replies from other scholars seeking to prove the efficiency of secured credit, but received significant support in a 1997 article by Lucien Bebchuk and Jesse Fried.[32] They showed that there are a substantial number of 'non-adjusting' creditors who are prejudiced by 'full' security given by the debtor to a secured creditor and who cannot or are unlikely to protect themselves against the higher risks posed by the security.

My point is not to resume here the debate over the efficiency of secured credit but to draw attention to its relevance to other questions, also under current discussion. These involve the goals of modern commercial insolvency law and, so far as England and the recently adopted amendments to the Insolvency Act are concerned,[33] the significant scaling back of a floating charge holder's right to veto administration proceedings,[34] and the earmarking of a share of the insolvent estate for the satisfaction of non-priority unsecured creditors' claims.[35] These questions pre-eminently lend themselves to scholarly analysis and discussion, particularly if (as I believe they should be) they are supported by relevant empirical studies. As if this cocktail of competing concepts were not enough, there is the further radical proposition being advanced by Prof Schwartz and other North American law and economics scholars that debtors should be allowed to contract out of some of the procedural protections afforded them under the US Bankruptcy Code so as to enable secured creditors to realize their security much more quickly than is currently the case.[36] Until recently, this normative analysis of insolvency law received very little attention in Canada or the United Kingdom. One of the many virtues of Vanessa Finch's recently published and most admirable book[37] is that she has squarely put these and many other policy issues in the forefront of contemporary analysis of insolvency law, not just for England but as well for other Commonwealth countries with advanced economies.

Another American aspect of the debate over bankrupt policies has, I believe, also significance for other common law jurisdictions. US commercial law scholars have not confined their analyses to learned books and

[32] 'The Uneasy Case for the Priority of Secured Claims in Bankruptcy: Further Thoughts and A Reply to the Critics' (1997), 82 *Cornell Law Review* 1279.
[33] See *Enterprise Act 2002*, c 40, Part 10.
[34] *Ibid*, s 250.
[35] *Ibid*, s 252.
[36] For a summary of these and other proposals for the revision of ch 11 of the Code, see Chas J Tabb, 'The Future of ch 11' (1993) 44 *South Carolina Law Review* 791.
[37] Vanessa Finch, *Corporate Insolvency Law. Perspectives and Principles* (Cambridge, Cambridge UP, 2002).

articles. Frequently they also make common cause to make written and oral submissions to Congressional committees opposing what they perceive to be ill conceived legislative initiatives. This happened several years ago with respect to credit industry sponsored efforts to reverse the hundred-year-old US 'fresh start' policy for consumer bankrupts.[38] It happened again more recently in relation to another industry sponsored draft amendment to the US Bankruptcy Code,[39] one whose goal was to make securitisation of assets bankruptcy proof. Thirty-five American insolvency law academics signed a letter to the chairs of the Congressional committees strongly opposing the amendments as mischievous and very damaging to basic bankruptcy concepts.[40] There are basic structural differences between the US governmental system and its division of powers between the president, the Congress and the courts, and the Canadian and British Parliamentary systems, which may provide fewer opportunities and less need for academic interventions of this type. Nevertheless, it would be wrong to dismiss the important political role commercial law academics in Commonwealth countries can play in *opposing* as well as advocating proposals for change. Certainly this is true in Canada where, at both the federal and provincial levels, legislation of a technical character is often passed with minimal scrutiny by legislatures.[41] In such cases, academic voices can surely play a significant role in drawing attention to legislative defects even if the criticism has no immediate impact.

VII. HARMONISATION OF INTERNATIONAL AND EUROPEAN COMMERCIAL LAW

Until comparatively recently, the efforts of commercial law scholars were largely confined to the domestic area. I have already noted how much this has changed with the growth of world trade, the establishment of the European Union and free trade areas elsewhere, and the proliferation of legal institutions, international conventions, model international laws, and industry inspired uniform documents and rules to promote the flow of commerce. It hardly needs stressing that common law legal academics have played as prominent a role in many of these initiatives as they do at the domestic level. Names such as Roy Goode, Hugh Beale, Jack Beatson, Michael Bridge in the United

[38] See C J Tabb, 'The Death of Consumer Bankruptcy in the United States?' (2001) 18 *Bankruptcy Developments Journal* 1.
[39] Viz s 912 in s 420/H R 433, 97th Cong (2001).
[40] See letter by the academics dated 23 January 2002 to Sen Patrick Leahy and Cong F James Sensenbrenner. The letter is accessible on various websites.
[41] A good example, at the provincial level, is the Ontario Consumer Protection Statute Law Amendment Act 2002, which involved a massive 200 page revision of Ontario's consumer protection legislation. The bill received several hours of debate during its second reading but was rammed through in committee in one day and no witnesses were allowed to appear before the committee.

Kingdom, Ron Cuming, Vaughan Black, and Catherine Walsh in Canada, and Allan Farnsworth, Neil Cohen, Jay Westbrook, and Peter Winship from the USA come readily to mind.

My point however is less to stress their individual roles—important as they are—than to reflect on an overriding theme that has engaged the minds of so many able scholars. This is the question of how harmonisation of national and regional commercial laws can best be promoted. There is general consensus about the different approaches available for this purpose, ranging from convergence of judicial decisions and domestic legislation at the bottom rung of the ladder to an international or European *lex mercatoria* at the top. It is the latter goal that has provoked the greatest controversy and has generated an enormous literature. It seems there are at least three different schools of thought. At one end are the optimists, who basically see no reason why enlightened lawyers should not be able to agree on common principles of international commercial law despite the lawyers' different systemic backgrounds. They point to such successes as the Convention on Contracts for the International Sale of Goods (CISG), the *Unidroit* Principles of International Commercial Contracts (*Unidroit* Principles), and the EU sponsored Principles of European Contract Law (European Principles) as proof of what has been accomplished. At the opposite end of the spectrum are those observers who are deeply sceptical and outrightly hostile to the suggestion that differences between deeply rooted legal systems can be wafted away by the adoption of seemingly inoffensive principles. Professor Legrand is probably the most articulate member of this group[42] but he is far from alone.[43] The suggestion that the civilian doctrine of good faith can, or should, be imported into English law has served as a particular lightning rod for these sceptics,[44] despite the fact that the good faith principle has long been accepted in American contract law and has also gained a significant foothold in Canadian law.[45] The members of the third school adopt an intermediate position. They support the crafting of model laws and transnational contract principles so long as lawyers and tribunals are free to adopt or reject their application to particular cases as they deem appropriate.[46]

[42] P Legrand, 'Against a European Civil Code' (1997) 60 *Modern Law Review* 44.

[43] For England see in particular, Sir John Hobhouse, 'International Conventions and Commercial Law: The Pursuit of Uniformity' (1990) 106 *Law Quarterly Review* 530.

[44] See G Teubner, 'Legal Irritants: Good Faith in British Law or How Unifying Law Ends Up in New Divergences' (1998) 61 *Modern Law Review* 11.

[45] See among others S O'Byrne, 'Good Faith in Contractual Performance: Recent Developments' (1995) 74 *Canadian Bar Review* 70 and *Wallace v United Grain Growers Ltd* [1997] 3 SCR 701.

[46] For my money, Prof Teubner, above, provides the most nuanced interpretation of this approach. Among commercial law scholars, Roy Goode is the most prominent member of this school. See Roy Goode, 'Reflections on the Harmonization of Commercial Law' in Cranston & Goode, *Commercial and Consumer Law. National and International Dimensions* (Oxford, Clarendon Press, 1993), ch 1. Prof Goode also supports UK ratification of the International Sales Convention and presumably does so because, for the most part, the parties are free to exclude the Convention. See Convention, Art 6.

My own position in this important debate is unimportant. What does seem to me important is that lawyers should conduct the debate realistically. Differences between legal systems should not be exaggerated, particularly when the same principles are applied, albeit in different guises. By the same token, apparent successes should not be inflated either. A significant number of UNCITRAL and *Unidroit* initiatives have only attracted limited state support and in some cases have not yet come into force.[47] Despite the 62 states that have ratified the International Sales Convention, the fact remains that major trading nations such as Japan, India and the United Kingdom have not yet done so. Still more important, the case law in the common law jurisdictions that have adopted the Convention—Australia, New Zealand, Canada, and the US, to mention the most prominent—is negligible and suggests that litigators in those countries avoid pleading the Convention because they find it more trouble than it is worth even when the Convention has not been expressly excluded in the parties' contract.[48]

VIII. THE NEED FOR NEW METHODS AND NEW APPROACHES

It is open for consideration whether the traditional methods for the analysis of legal issues have ever been adequate over the past fifty years, not to mention earlier periods. Too often when reading even a very scholarly text one is left with the impression that contemporary law consists of a long string of cases, interspersed with an increasingly heavy freight of statutory provisions, and too little discussion of principle to test or illuminate the rationale of a rule or doctrine, or its adequacy in a rapidly changing economic and social environment. To be sure, the criticism may be unfair and may fail to recognise that courts must decide, lawyers must advise, and government ministers must take a position, and that none of them can afford the luxury of indeterminate theorising about the ideal rule. Moreover, there are many excellent articles and judgments on both sides of the Atlantic in which authors and courts supply the missing links for which there may not be sufficient room in a standard textbook.

Nevertheless, in my view, the normative process can be substantially enriched by making better use of the insights of non-legal disciplines, especially (in the case of commercial law) the insights of the economic analyses of law, and, by using empirical investigations, determining how the law

[47] Report of the United Nations Commission on International Trade Law on its thirty-fifth session, 17–28 June 2002, XIV. Status and Promotion of UNCITRAL Legal Texts, pp 41–2.

[48] See JS Ziegel, 'The Future of the International Sales Convention from a Common Law Perspective' (2000) 6 *New Zealand Business Law Quarterly* 336 and, for an American perspective, MW Gordon, 'Some Thoughts on the Receptiveness of Contract Rules in the CISG and Unidroit Principles as Reflected in One State's (Florida's) Experience of (1) Law School Faculty, (2) Members of the Bar with an International Practice, and (3) Judges' (1998) 46 *American Journal of Comparative Law* 361.

operates in practice, and encouraging closer contacts between legal scholars, the judiciary, and practising members of the professions. Some observations are appropriate about each of these themes.

A. Economic Analysis of Law

Commercial law is fundamentally about markets and profits, enforcement of contracts, and remedies for breach. One would therefore expect commercial law teachers to show great interest in the economic analysis of law because it addresses the same issues. However, this does not appear to be true in the UK. The 2002 *SLS Directory* only lists six members with particular interests in the area. Leading texts either do not mention the economic analysis of law at all or give it very little space.[49] However, it may be that these numbers do not give a fair impression of what is happening in the UK and that they fail to take into consideration some of the more recent developments.[50]

In any event, the position is very different in the US, the birthplace of the law and economics movement. Here the economic analysis of law is well entrenched at all the leading law schools. In the 1980s, sceptical commentators predicted that the movement would peak and lose momentum[51] but the opposite appears to be true.[52] The discipline has generated a huge literature, much of it of high quality, and has spawned many specialist journals.[53]

[49] GH Treitel, *The Law of Contract,* 10th edn (1999) does not discuss it nor, in Canada, does SM Waddams, *The Law of Contracts,* 4th edn (1999). Roy Goode, *Commercial Law,* 2nd edn (1995), pp 25–26, only devotes two pages to the topic, although he recognizes that 'much valuable work has been done ... in testing principles and rules of commercial law against concepts of economic rationality and efficiency and in challenging assumptions commonly made by lawyers as to the appropriate approach to the resolution of particular issues.'

[50] My colleague Kevin Davis has drawn my attention to the important work being done by the newly established ESR Centre for Business Research (CBR) at Cambridge University. However, I have not been able to locate reports that have a direct bearing on commercial law. Tony Ogus has also suggested to me that while commercial law scholars may so far have shown little interest in law and economics in the UK, the position is different in the company law area. He did not provide details. I am familiar of course with Brian Cheffins' excellent text Brian R Cheffins, *Company Law: Theory, Structure and Operation* (Oxford, Clarendon Press, 1997), but I don't know how far it has changed the teaching and research methods of company law teachers outside Cambridge University.

[51] Cf R Posner, *Economic Analysis of Law,* 5th edn (Aspen, 1998), p viii, n 1, citing MJ Horwitz (1980) 8 *Hofstra Law Review* 905 and Owen M Fiss (1989) 74 *Cornell Law Review* 245.

[52] Prof Langbein assures us in JH Langbein, 'Scholarly and Professional Objectives in Legal Education: American Trends and English Comparisons' in Birks, *What Are Law Schools For*?, v 2, 1 at 4, that 'If there is one message that I could leave with an audience of English legal academics, it would be to emphasize the fundamental and revolutionary importance of law-and-economics, not only for American legal education, but for the shape and character of American law.' He also notes that half of US sitting federal judges have undergone law and economics training courses (*ibid*).

[53] The University of Toronto law school library catalogue lists 18 law and economics journals. However, some of them are published in Europe.

There is also a strong American Law and Economics Association whose meetings are well attended.

Canadian law school interest in the economic analysis of law is not as strong as in the US but appears to be greater than in the UK. This is particularly true of the University of Toronto law school, which has a very active law and economics programme and at least seven faculty members with major interests in the area. The programme includes a law and economics seminar, held throughout the calendar year, at which papers are presented by visiting scholars and faculty members. The law school also hosts the annual meeting of the Canadian Law and Economics Association (CLEA). Interest in the economic analysis of law is scattered at other Canadian law schools and is probably only marginally greater than it is in the UK. Similarly, outside the competition law area, Canadian courts have until recently paid little attention to the economic analysis of legal issues but there are signs that this may be changing.[54]

How do we explain the general lack of interest in the economic analysis of law in Canada and the UK outside the special cases of Cambridge and Toronto? I have put the question to English and Canadian colleagues. Coupled with my own impressions, the answers appear to be the following. Many law teachers lack an economics background and feel discouraged by the technical jargon and the use of tables and graphs that appears in much law and economics literature, not to mention the use of calculus and regression tests! Articles are often written very abstractly and seem to bear little relation to facts encountered in actual cases. Too many lawyer economists are perceived to embrace right wing ideologies and to be committed to free market solutions to the exclusion of other values. Too often markets are idealised and insufficient allowance is made for very imperfect markets and informational asymmetries. Law and economics' critics also express scepticism about the basic premise that human actors always behave rationally to maximise their utilities.[55]

I do not discount the importance of some of these reactions. Nevertheless, I also believe the economic analysis of legal issues has too much to contribute to a better understanding of the legal system to deserve to be ignored. Lack of prior knowledge of economics can be addressed (as it is addressed at a number of Canadian law schools) by offering students an introductory course in the subject. Many fine lawyer economists have

[54] See for example Major J's judgment for the Supreme Court of Canada in *Bank of America Canada v Mutual Trust Co* (2002) 211 DLR (4th) 385 (SCC) employing economic reasoning to justify awarding the plaintiff compound interest and distinguishing the efficient breach rule.
[55] A new branch of economics, known as behavioral economics, has developed to address these questions and researchers have made important findings. See CR Sunstein, (ed), *Behavioral Law and Economics* (Cambridge, Cambridge UP, 2000). The frenzied share dealing in new technology companies leading to the collapse of the dot.com securities market in Europe and the US in 2000 and the vast overbuilding of fibre optic cable networks in North America and Europe are recent illustrations of such irrational behaviour.

no honour degrees in economics, and there are many excellent articles and books written in non-mathematical terms that are quite intelligible to a lawyer willing to take the trouble to read them.[56] I venture to suggest that the most promising approach in Canada and the UK is to develop centres of legal excellence with strong graduate programmes that will include law and economics subjects in their curricula but not, of course, to the exclusion of other more traditional subjects.[57]

B. The Importance of Empirical Studies

In the common law world, Roscoe Pound drew attention at the turn of the century to the difference between black letter rules and the law in practice.[58] The same theme was picked up by the legal realists in the 1920s and 1930s although they sometimes used it to pursue their own idiosyncratic interests. Commercial lawyers need not be reminded of the importance of Roscoe Pound's message, but it is disappointing to see how little empirical work has been done in England and Canada on how markets and legal rules work in practice.

We know for example that the mirror image rule governing offer and acceptance is often ignored in the exchange of standard form contracts. We do not know, however, how pervasive the practice is and whether it means that the rule itself needs to be changed.[59] We know too that commercial contracts for the sale of machinery, vehicles and other high priced durable goods regularly exclude the manufacturer's liability for consequential damages if the goods turn out to be defective and that they restrict the buyer's right to reject non-conforming goods.[60] Here again, little empirical work appears to have been done on the practical effect of such disclaimer clauses, whether they represent a more efficient allocation of risks between manufacturer and buyer than the rules in *Hadley v Baxendale*,[61] and if they do

[56] An eminently readable text is that by R Cooter and T Ulen, *Law and Economics* (Scott, Foresman & Co, 1988).

[57] In Canada, as in the US, law is treated as a graduate programme and students are generally required to have a first degree in another discipline as a condition of admission to the law school. In this way, many of the students will come to law with some prior exposure to economics. A substantial percentage of law students also have graduate degrees. In Canada, a growing number of University departments (including the University of Toronto law school) also offer joint four year MBA/LLB programmes, which again ensure that the students receive strong doses of applied and pure economics.

[58] R Pound, 'Law in Books and Law in Action' (1910) 44 *American Law Review* 12.

[59] As is well known, s 2–207 of the American Uniform Commercial Code attempted to do just this but with very unsatisfactory results. Cf Ontario Law Reform Commission, *Report on Sale of Goods* (1979) v 1, pp 81–85 ('OLRC Report').

[60] OLRC Report, v 2, pp 461–63, 485–89. A substantial number of provincial statutes avoid the use of disclaimer clauses in consumer contracts but these provisions do not apply to commercial buyers.

[61] (1854) 9 Exch 341.

whether this suggests the need to amend the existing default rules in the Canadian and British Sale of Goods legislation.

To give a further example, in the insolvency area, there is an ongoing debate about the relative merits of different statutory schemes for rescuing insolvent companies. Despite the importance of the question, many of the existing rules and proposed legislative changes are based on assumptions unsupported, or not adequately supported, by hard data. The same observations could be made about many other aspects of Anglo-Canadian insolvency law.

In the pure and applied sciences, theories and hypotheses are expected to be verified by experiments and actual observations. Puzzlingly, the same need does not appear to be felt in most branches of the common law, including commercial law. There may be a variety of reasons for this difference[62]—lack of funds, lack of incentives and prestige in doing empirical work, and lack of experience in the design of questionnaires and use of statistical techniques, to name some of them. Nevertheless, I am not persuaded the difficulties are sufficiently great to justify our ignoring potential gold mines of information to gain a better understanding of how legal systems and legal rules actually operate at the ground level.[63]

C. Cooperation with Law Firms, Judges and the Bar

It is a great conceit to suggest that law teachers working in an academic environment have a lockhold on scholarly study of commercial law issues and I hope this chapter does not leave this impression. My intention is quite the reverse. Law firms, barristers and judges are great repositories of learning and experience and possess superb skills that legal academics should be anxious to tap for the common good. Happily, this symbiotic relationship between the legal academy and other branches of the profession has greatly accelerated over the past twenty-five years. In both England and Canada, practitioners frequently assist in seminars and offer specialized courses at law schools; in both countries major law firms have endowed chairs in law and frequently sponsor conferences and colloquia. Legal scholars are quite

[62] Some of them are explored in Jay L Westbrook, 'Empirical Research in Consumer Bankruptcy' in *Symposium: What We Know and Do Not Know about the Impact of Civil Justice on the American Economy and Polity*' (2002) 80 *Texas Law Review* 2123.

[63] In the US a great deal of empirical work has been done, and is being done, on many aspects of consumer and commercial bankruptcies, notably by Profs Jay Westbrook and Terry Sullivan at the University of Texas and by Prof Elizabeth Warren at Harvard. (The three often work as a team.) See also Symposium, 'Empirical Research in Commercial Transactions' (2000) 98 *Michigan Law Review* 2421. In England, Sally Wheeler wrote an excellent doctoral thesis exploring the practical impact of the *Romalpa* clause but her findings seem to have attracted little attention. See S Wheeler, *Reservation of title clauses: impact and implications* (Oxford, Clarendon Press, 1991).

often also retained by law firms to advise on complex problems or to run in-house educational programmes for members of the firm. Obviously, there are risks, as well as opportunities, in legal scholars working closely with law firms. By giving advice, especially in sensitive ethical or otherwise controversial areas involving basic common law norms, the academic runs the risk of compromising his independence.[64] Nevertheless, I do not believe this is a sufficient reason for the scholar to withdraw into his academic shell. It does, in my view, call for good judgment by the academic and careful evaluation whether the job for which he is being retained is compatible with his future objectivity and independence as a scholar.

So far as linkages between the judiciary and law faculties are concerned, I have previously referred to the substantial number of academics in Canada who have been appointed to the bench, and have noted the encouraging signs of the beginning of a movement in the same direction in England. In both countries, judges often participate in academic conferences and, in England at least, often deliver papers and special lectures.[65] In Canada, ties between the courts and law schools have also been greatly strengthened since the late 1970s through the appointment of highly qualified recent graduates to serve as law clerks to the Supreme Court of Canada, the Federal Court of Canada, and provincial appellate and superior courts.[66] This cross-fertilisation of resources has undoubtedly greatly benefited both the academic community and the courts. I would expect similar happy results from the recent introduction in England of the hiring of young lawyers to serve as legal assistants in the Privy Council and the House of Lords.[67]

There has been a further innovation in Canada which may be of interest to judicial authorities in other common law jurisdictions. This involves giving a small number of selected judges each year study leave for a period of up to seven months to enable them to pursue an approved programme of study, research or teaching (or a combination of two or more of these), usually at a Canadian law school.[68] The programme has been very successful

[64] The same is true of course of the solicitor-client relationship and the *Enron* debacle furnished striking examples of the latter phenomenon.
[65] Encouragingly, members of the English Court of Appeal played active roles at the LSE conference at which this paper was presented.
[66] See above n 10 and accompanying text.
[67] See notice of 12 March 2002 by the Judicial Office of the House of Lords sent to firms of solicitors. (I am indebted to Prof Ross Cranston, MP, for providing me with an email attachment of the notice.)
[68] Some of the other particulars are that the candidates must have served at least seven years on the bench and their applications must be approved by the chief justice of the court to which they belong. Applicants must submit a programme of study and the successful candidates are selected by a joint committee of law deans and the Canadian Judicial Council Successful candidates also require the approval of an order in Council from the federal government. (Under Canada's constitution the central government is responsible for the appointment of superior court judges to provincial as well as federal courts.) (I am indebted to Justice Robert Sharpe of the Ontario Court of Appeal for much of this information.)

so far and has paid rich dividends for both faculties, students and judges. It is not difficult to conceive how greatly enriched English law faculties would be as well if the British government could be persuaded to introduce a similar study leave programme for members of the High Court and the Court of Appeal. To be sure, commercial law scholars would be only one of the beneficiaries. However, given the importance of commercial law to the British economy and increased pressure for integration of the legal systems of the member countries of the EU, even a share for these scholar of this new resource would be no mean gain.

My suggestion may seem far-fetched to English readers. They may find it inconceivable that the Treasury would support such expensive use of judicial talent and also query how many judges would be interested to spend a leave of absence in an academic environment. The first of these objections is obviously the weightier one but, in my view, it could be overcome by the British government's willingness to earmark a percentage of High Court and county court registry fees for supporting judicial leaves. Another possibility would be for law schools to establish judicial fellowships, perhaps with the financial assistance of charitable endowments that have been active in financing other legal visitorships in the past.[69]

IX. ENVOI

I have presented an optimistic picture of the present and future prospects for commercial law scholars in England and Canada. However, it would be a mistake to suggest that all is plain sailing and that there are no obstacles to overcome to enable the academy community to play its optimal roles.

In fact, there are several such obstacles and they merit brief discussion. One was mentioned earlier[70] and involves the difficulty of law faculties recruiting as teachers and retaining the best graduating students given the much higher salaries available in practice and the high cost of housing in major urban centres in both England and Canada.[71] The recent tuition increases implemented at several Canadian law schools[72] should enable them to offer substantially better salaries in the future. Presumably the same result will flow from the British government's decision in January 2003 to set British universities free to determine their own tuition fees up to defined limits.[73]

[69] I note with interest that Ross Cranston, MP, is a professorial fellow at the LSE and no one would accuse him of shirking his Parliamentary duties!
[70] Above text accompanying n 8.
[71] In repeating this observation I do not mean to suggest there is a shortage of applicants for teaching posts at Canadian law schools; there is not. The shortage arises in attracting the best candidates and I understand the same is true in England.
[72] Above n 8.
[73] *The Economist*, 25 January 2003, p 16 (N American edition).

These changes are to be welcomed. Nevertheless, it is inevitable that there will always be a large gap between the incomes of successful barristers and solicitors and the salary that even the holder of a prestigious chair at a senior Anglo-Canadian law school can hope to command. We have come to accept such disparities on the footing that law teachers choose an academic career because of the satisfactions it offers and not because of the material rewards. However, there is a limit to the financial sacrifices that even the best motivated academics can be expected to make for themselves and their families.

An equally important challenge facing the modern law school with serious research aspirations is the development and maintenance of adequate library facilities. According to John Wilson's findings,[74] based on a 1991 survey, British university law libraries do not fare well in this area and the great majority of them at the time had holdings of less than 50,000 books. By way of contrast, the average holdings of Canadian university law libraries is over 200,000 volumes.[75] Inadequate library resources make it difficult for a commercial law scholar to pursue serious research unless he chooses his topic judiciously.

Admittedly, access to WestLaw, LexisNexis and other electronic sources have provided a bonanza to legal researchers on both sides of the Atlantic. However, there are still many commercially sponsored journals and reports that are not available electronically or only available for a hefty subscription price. The same is true of many if not most textbooks and monographs, which are often only available in paper form and are often very expensive. The constraints on library budgets and the non-availability of research materials is most keenly felt when the scholar attempts to do comparative work. It is probably fair to say that this is why many British scholars find it difficult to follow commercial law developments in Canada and other Commonwealth jurisdictions (and *a fortiori* US developments) and why Canadian scholars are similarly handicapped in following the British scene.[76]

[74] 'A Third Survey of University Legal Education in the United Kingdom' (1993) 13 *Legal Studies* 143, 163, Table 19.

[75] A 1997 *Maclean Magazine* survey found that Osgoode Hall, Toronto and Alberta had the largest holdings (468,302, 423,825, and 389,010 respectively) and Moncton, New Brunswick and Saskatchewan the lowest at 120,019, 129,521 and 177,084 respectively. (Information kindly supplied to author by Shikha Sharma, reference librarian, University of Toronto law school.) I appreciate the comparison between Canadian and British holdings may be misleading because there are many more law reports in Canada than there are in the UK and they churn out many more volumes annually. Nevertheless, the comparison is meaningful in showing that Canadian law libraries enjoy significantly higher budgets than their British counterparts.

[76] I can illustrate this point from personal experience. Until I came to London for the LSE conference and visited the Lincoln's Inn Law Library I was not even aware of the existence of some English insolvency law journals even though they contained valuable materials of great importance to an overseas researcher trying to understand what is happening in the UK.

A serious academic scholar not only needs superior library facilities; he also needs a research budget for the purchase of books, to hire student research assistance, to pay for extra secretarial help and to attend conferences. The availability of research funds differs widely at Canadian law schools. Most of them provide some funding for the preparation of teaching materials (often with student assistance) and to enable the faculty member to attend at least one conference a year. Little is certain beyond this point and if additional resources are not available internally the Faculty member may have to approach an outside agency, such as the provincial law foundations in Canada and the Foundation for Legal Research of the Canadian Bar Association, in the case of small grants, and the federally based Social Sciences and Humanities Research Council for major grants. There is of course no guarantee that the applicant will be successful. I gather the funding position is substantially similar in the UK, which is small consolation to the researcher who has to spend valuable time beating the bush for financial assistance.

Finding a publisher for his scholarly output is another challenge facing the academic. In Canada there are reputed to be no less than ninety-nine law journals with academic aspirations[77] and finding a home for a well-crafted article or comment usually provides little difficulty. The difficulty arises in trying to persuade a Canadian publisher to publish a theoretical, academic or highly specialised monograph for which there is likely to be a very limited market. Commercial publishers will usually decline the invitation and even University presses will often only express interest if they feel confident the sales will be self-financing or that a publication grant can be obtained from an outside agency. As a result many Canadian legal scholars have found it much easier to publish their monographs with English publishers such as the Oxford University Press, the Cambridge University Press and, greatly to Richard Hart's credit, Hart Publishing in Oxford.[78] The reader will correctly infer that I believe English legal academics find a more receptive market for their manuscripts in England than do their Canadian colleagues in Canada.

I turn finally to the last of the obstacles facing modern commercial law scholars, one which, ironically, makes the other obstacles—indeed the whole commercial law enterprise—redundant. In 1992 Judge Edwards, a federal appellate judge and himself a former member at several leading US law schools, published a much cited article[79] strongly criticising the leading

[77] This estimate was provided to me by Shikha Sharma, reference librarian at the University of Toronto law library, who arrived at the number by perusing the Index of Canadian Legal Periodicals and culling from it what she believed to be journals with appropriate qualifications.
[78] Astonishingly, Hart Publishing has managed to publish 300 titles since the company was founded in 1996. Would that there were a Canadian counterpart!
[79] See Harry T Edwards, 'The Growing Disjunction between Legal Education and the Legal Profession' (1992) 91 *Michigan Law Review* 34.

law schools for having abandoned traditional legal scholarship and teaching in favour of the new 'law and' hybrid disciplines. As a result, he complained, the judiciary and the legal academic community no longer spoke the same language and courts could no longer look to the leading law journals for assistance in finding answers to current legal problems. Not surprisingly, US legal scholars promptly rejected Judge Edwards' criticisms[80] as based on an obsolete view of the role of the modern law school and the nature of the legal system.

Having quickly perused some recent volumes of the *Harvard Law Review* and the *Yale Law Journal,* I am not sure Judge Edwards' strictures are justified. The *Harvard Law Review* still contains excellent student notes on current legal issues and the annual review of US Supreme Court decisions, and I found little that disturbed me about the character of most of the articles in both journals. In any event, there are hundreds of US law journals and, so far as I can judge, no shortage of articles dealing with doctrinal materials. There is even less danger, in my view, of Canadian and English scholars being seduced by the alleged US law schools' obsession with interdisciplinary approaches to teaching and scholarship. Doctrinal legal scholarship is alive and well in the UK and Canada as is attested by many of the contributions to this volume and as may be seen by consulting recent volumes of the *Modern Law Review*, the *Law Quarterly Review*, the *Canadian Bar Review* and other leading Anglo-Canadian journals. There may be more articles and more research of a theoretical, empirical and interdisciplinary character than was true twenty-five years ago and I suggested earlier that this was a desirable development. I do not believe it is a cause for alarm and do not anticipate that it will in any way impair the cordial relationships that have been so successfully nurtured over the past decade or more in England between the academic community on the one hand and the judges and the legal profession on the other.

[80] See 'Symposium, Legal Education (1993) 91 *Michigan Law Review* 1921 and compare with Langbein, above n 52, supporting Judge Edwards' perceptions.

Commentary on 'The Legal Academy's Contribution to the Development of Commercial Law: An Anglo-Canadian Perspective'

Legal Academics and Legal Practitioners—Are They on the Same Planet?

TONY KING

I. THE KEY THEME TO THIS COMMENTARY

WHILE THE TITLE to this commentary may appear provocative, I chose it with care. It does reflect a current, but inaccurate view of academics and practitioners, one which can colour the relationship between the two sides. In my view, it must be consigned to the waste bin of history for the benefit of all. So, what planet are academics and practitioners on? The same one, they just don't see it in quite the same way!

All 'lawyers' (to use an all-encompassing generic term) look at one thing—the law. In some instances, academics and practitioners look at some legal issues in precisely the same way, but they bring different experiences to the situation. In other situations, they may look at the same thing from different perspectives. In either case, one side not drawing on the knowledge and experience of the other will mean a missed opportunity for an enriched analysis. Academics and practitioners using each others' expertise in appropriate ways can only lead to a more effective analysis, development and application of the law. This is a situation where the sum of the whole will definitely be greater than its parts.

This happens already. However, in this commentary I will advocate that academics and practitioners should work ever more closely together. Through this symbiotic relationship the maximum use will be made of the very considerable skills, experience and expertise which reside in 'lawyers' as a group. This is not to imply that the academic and practical approaches should merge into a single approach—the value is in maintaining the differences but ensuring greater interchange between the two sides.

II. CAVEATS

The views I have expressed in this paper are my own; I am not purporting to express the views of my firm and this paper should be read in that light. The ideas are based on more than 30 years of exposure to the law as:

— a student;
— an articled clerk (as trainee solicitors were called);
— a law teacher;
— a practitioner;
— an in-house trainer with a law firm.

I have not embarked on a major research project the results of which I am drawing on to justify the views of this commentary. However, I am drawing on three decades of experience of the law and of lawyers (probably now numbering in the thousands) working or studying in the law's many branches—in law faculties and professional teaching institutions, in law firms big and small, English or international, working in-house or in private practice, at the Bar or in the judiciary.

While I have worked with lawyers from across the globe, my principal experience of the law is in England and Wales. This paper focuses on my experiences in this country and I do not attempt a broader international analysis or comparison.

This is intended to be an accompanying commentary on the chapter entitled 'Legal Academy's Contribution to the Development of Commercial Law: an Anglo-Canadian Perspective' by Professor Jacob Ziegel. That paper is an excellent analysis of the benefits of a closer relationship between academics and practitioners. This paper will pick up on a number of the thoughts and proposals which are expressed so clearly by Professor Ziegel.

III. WHY IS THERE A PERCEIVED DIVIDE BETWEEN LEGAL ACADEMICS AND LEGAL PRACTITIONERS?

Before illustrating today's close working relationships between academics and practitioners and the opportunities for mutually valuable future cooperation, it is worth giving some thought to why there is a perception, real or otherwise, of a divide between the two sides.

I am wary of looking for possible explanations for this perceived divide as any explanation runs the risk of causing someone offence. That is most certainly not my intention. Equally, by asking why this perception exists I could be seen as advocating merging the roles of academic and practitioner. I am not. I feel the differences should be respected and capitalised on, not despised or derided. In a paper such as this, I have no choice but to generalise and, of course, all generalisations can be refuted by specifics. However, whatever the rights and wrongs of the views I have set out in this section, I have heard all of them expressed at some point in my career.

A. The Study of Law is Different from its Practice

No one, including me, would deny that the skills and expertise which legal academics and practitioners bring to their respective roles are different. However, they are not completely different. Both involve research skills, analytical and problem solving skills, communication and interpersonal skills. The 'raw material' (the law) in respect of which these skills are exercised is often the same, but although the outcome of or purpose for the application of those skills can be different, the differences are not so great that it is justifiable to regard academics and practitioners as operating in different intellectual worlds.

Let's look at the different outcomes. A practitioner may analyse the sources to find what the law is now on a particular topic with a view to advising a client appropriately. An academic may go through precisely the same process to explain what the law is to his or her students. The process underpinning the explanations will be the same, it is the focus which is different.

The very nature of the study of law at the highest levels may enable an academic to research widely around the topic to identify not only what the law is but also to analyse how it came to be what it is and make suggestions for what it could or should be. The commercial pressures which many practitioners face may mean that the opportunities to act as such a catalyst for change are taken up less often. However, many practitioners do undertake (often with their academic colleagues) to lobby for or advise on much needed improvements in the law, both as a matter of public good, and in the interests of their clients. Some will regard this as just as much a part of their normal practice as more conventional work. There is very considerable scholarship undertaken by practitioners and many academics make significant contributions either direct or indirect to the practice of law.

A different approach does not mean a fundamentally different role.

B. You Study Law as a Student, then You Get on with Real Life Practice

This statement can be interpreted in a number of ways. It can indicate a belief that lawyers have no need to continue learning once they have gained their academic and professional qualifications. This attitude overlooks the necessity for all lawyers to maintain and update their legal expertise. Study may take different forms and be prompted by different drivers but, to take the practitioner angle, one senior partner said to me at his retirement party, 'I stopped learning the law this afternoon—the last day of my practising career'. In reality, the view of this partner is the one which I would expect all practitioners to hold. However, some young lawyers may have a different view, believing that learning has ended as they make the transition from being students to starting their professional careers.

Resisting the temptation to embark on psychological theorising, in my view this is as a result of a tendency to compartmentalise. Many young would-be lawyers are very target-oriented and task-focused. They set themselves the task of gaining good grades at school as a way of ensuring they get into the best university they can. Once there, again they work hard to achieve the best grades with a view to getting themselves on the first rung of the best career ladder they can. All this is understandable but it can mean that they treat their law studies as being an extension of their school days, in other words something one does when young before embarking on a career. That can lead to the study (and, by implication, the teaching) of law being seen as irrelevant to practice. This is clearly mistaken, but it can happen.

Another more serious consequence of the compartmentalisation is the risk that students do not fully understand the importance of their law studies as a foundation for their long-term careers. This means that, however interested they may be in the subject and however well taught, they may see the course as being merely one step on the road to gaining a good degree (their current goal). It may be reasonable to regard some school studies as being of limited use in later life—I for one have not used calculus since I was sixteen. However, contract law is at the heart of most practitioners' careers.

Obviously, not every student compartmentalises their studies in this way. I am nevertheless surprised by the number of students who ritually destroy their course notes once they have passed the exam. This compartmentalisation can lead to young lawyers seeing the various steps on their career as being like stepping stones, separate and unconnected from each other, rather than a continuous road. A consequence of this is that some students see their academic legal studies as in some way separate or distinct from their lives as practitioners, not the foundation of a future career. While I do, of course, accept that this is not necessarily at the top of the priority list for every undergraduate as they go through their studies, the attitude is regrettable especially in view of the very considerable efforts of academic lawyers to put the law into context.

C. Academics Teach Law, Practitioners Teach Practice

This is positive in that it reflects a general acceptance by practitioners that a career in law involves 'life long learning'. But it is unfairly dismissive of what academics do. So, how does this attitude come about? English solicitors, like many professionals around the world, are subject to the Law Society's Continuing Professional Development scheme (requiring at least 16 hours of continuing development each year, achieved through various methods). While this may be a lever for encouraging continuing study by practitioners, many voluntarily exceed the Law Society's minimum requirement by a considerable margin. Needless to say, in addition to any formal training enormous amounts of learning happens on the job through working for clients, coaching, mentoring and so on. Therefore (and as is only to be expected), the bulk of the day-to-day teaching a young practitioner receives will be from his or her more senior practitioner colleagues. This, coupled with attendance on the wide range of formal training opportunities whether in the form of in-house or public programmes (often, but certainly not always, led by practitioners), can lead to a perception that fellow practitioners are the main, if not the only, source of practical training.

If this leads to the perception that academics focus on 'pure law' and practitioners on 'practical law', it could be a natural and appropriate consequence of the purposive differences in their roles. However, it does not and should not mean that each of those areas is the exclusive preserve of one side or the other. From the outside, this apparent division does not seem to apply to, for example, the medical profession so why does it seem to apply to law?

D. The Specialist Bar is there to Help with Difficult Areas of the Law, not Academia

In many jurisdictions, as Professor Ziegel explains in his paper, academics are called upon by practitioners for help with interpreting the law. While that does, of course, happen in this country, many solicitors would look to the specialist Bar for this help.

To risk generalising again, the nature of practice at the specialist Bar involves a very considerable academic analysis of the practical problems the practitioners face. Many practitioners, therefore, turn to the specialist Bar for much needed advice and guidance presented in their own terms. Perhaps the explanation is that the practitioners see the specialist Bar as engaged in the pure study of the practical application of law.

While many academics have successful practices at the Bar, I am not suggesting that an expansion of this is necessarily desirable. What is desirable is a greater recognition that the views of academics are as valid on many practical issues as those of practitioners.

E. Allowing Non-law Graduates to Qualify as Practitioners Undermines Academia

There has been very considerable debate about this issue. The opponents and supporters of this qualification route are each able to put forward strong arguments in support of their respective cases. The fact that this route is open does give practitioners access to a wider pool of extremely talented entrants to the profession. That in turn gives considerable assistance with helping the practitioners manage their personnel planning requirements. It is true that a non-law graduate will have had less exposure to the academic study of law by the time they start their practising careers. They will have studied a narrower range of legal topics and had less time to acquire or hone and practice their 'lawyerly' skills. On the other hand, most, if not all, university disciplines require their under-graduates to develop the generic skills of research, analysis, problem solving and communication. It is my experience that if non-law graduates are at any disadvantage at the start of their Training Contracts, that disadvantage has been addressed by the end of the two-year period.

Whatever the merits of this qualification route, it does encourage wider access to the profession and the law is not alone in taking this approach. Accountancy and medicine do the same.

F. Is this Perceived Divide a Problem?

In this section, I have tried to identify the common 'prejudices' or causes of tension between academics and practitioners. I may not have identified them all and some of my comments can be disputed. However, for good or for bad I trust they go someway to explaining why academics and practitioners can be seen as separate and distinct.

The key issue is whether there is a real barrier between academics and practitioners. My view is that what may be seen as a barrier is merely a difference. This is something which should be preserved so that the two sides can work together in their complementary ways to achieve a better end result.

IV. MAXIMISING THE BENEFITS THROUGH A SYMBIOTIC RELATIONSHIP

There are many instances of academics and practitioners bringing together their complementary skills to deliver outcomes which are better than each could have achieved separately. The success came because they were different but it also came from them working together either directly or indirectly. In this section, I want to look at some options for working together symbiotically. The instances I have covered are not especially innovative nor exhaustive and are certainly not intended to indicate that I believe there should be a fundamental change in approach by one or other side. They are merely instances of where complementary skills have been proved to have considerable benefits.

A. Training Programmes

Developing as a lawyer, whether as an academic or a practitioner, involves life long learning. The majority of lawyers spend their lives outside academia so, focussing on those practitioners, how do they gain and maintain their legal knowledge? Acknowledging this is something of an over-generalisation, in the early years, from academia; in the later years, from practice. The point is that a young lawyer acquires his or her legal knowledge in a continuous process incorporating formal study and practice. Therefore, close liaison between academics and practitioners can only benefit the development of lawyers.

There are many examples of this being done successfully at the undergraduate, professional training, traineeship and post-qualification stages. One illustration is the initiative of eight London firms in working together with three providers of the Legal Practice Course to develop, within Law Society guidelines, a programme which is the first year of, in effect, a three-year traineeship. Many faculties invite practitioners to give courses or act as guest speakers to give students an alternative insight; many firms rely very heavily on academics to meet the training needs of their qualified lawyers.

B. Research and Development

Very considerable research and development is undertaken both in practice and in academia. There is scope for more exchange of that effort, be it through secondments between academia and practice, sponsorship, joint conferences and publications. A key challenge is finding ways of bringing the right people together in the right way.

C. Developing the Law

Professor Ziegel has described a number of instances in which academics and practitioners have contributed to the development of law whether through empirical research or by the marriage of academic and practical view points on a particular problem. The drivers for these initiatives can vary—they may be the proactivity of a particular lawyer, the requirement of a government or regulatory body or even a practical problem faced by the public. Again, the challenge is bringing the right people together.

V. WHAT ARE THE CHALLENGES TO BE ADDRESSED IN ACHIEVING A TRUE SYMBIOTIC RELATIONSHIP?

Perhaps, and most importantly, we should encourage the recognition between academics and practitioners of the skills and expertise each possesses, and seek to show how they are complementary. Many lawyers are enlightened on this point already but more need to be persuaded. There are many very successful initiatives instigated by academics and practitioners which are achieving this aim but more needs to be done. We need to create more opportunities for each side to see the real contribution which the other could make. Are there enough fora for the exchange of ideas between academics and practitioners?

While there is no question that the work of academics and practitioners is equally valuable in the non-monetary sense, often practitioners are better resourced and better remunerated than academics. As Professor Ziegel explains in his chapter, this can put constraints on the research and other resources which are available to academics to the potential prejudice of their work. More worryingly, this can lead to a reduction in the number of academics. Of course, not everyone is motivated by money but it is understandable in times of spiralling student debt for young lawyers to put considerable weight on financial factors when making their long-term career choices.

This funding issue is potentially a significant challenge. There may be ways round it—academics could engage in practice or offer research and development services; practitioners could offer sponsorship in some form. By working together, perhaps academics and practitioners can find the most satisfactory solution.

VI. CONCLUSION

While by no means universally held and certainly rapidly changing, there is in our community a sense of a divide between academic and practitioner lawyers. By creating more opportunities for interchange between the two sides, this perceived divide should begin to disappear. Academics and practitioners should work together to find ways of achieving this while preserving their differences in approach and purpose.

Academic and practitioner lawyers are definitely on the same planet, they just look at things differently.

23

Contracts, Contract Law and Reasonable Expectations

ROBERT BRADGATE

I. INTRODUCTION

THE OBJECT OF the papers in this chapter is to explore the relationship between commercial law and commercial practice. Many of the papers focus on specific areas of commercial practice. The focus of this chapter is wider, on the relationship between contract law, and especially one familiar area of contract law, the rules concerned with what is termed 'contract formation', and legal and business practice. The objective is to explore the match, or mismatch, between contract doctrine, as found in decided cases and expounded in the books, and contracting practice. The focus is on commercial contracts in the wide sense of non-consumer business, as opposed to purely mercantile or large-scale commercial contracts. It will be suggested that the description of the law of contract in the traditional academic texts is a world away from the experience of contracts in real life; that bald statements of the law in texts, with an excessive emphasis on doctrinal certainty, may sometimes misdescribe the reality of the law in application and, in the process help generate false expectations; and that the rules themselves are often at odds with business practice. In short, the books misdescribe the law and, possibly as a result, the law on occasion is misaligned with the expectations of contractors. Lord Steyn has suggested that the touchstone of the law of contract should be that it protects or upholds the 'reasonable expectations of honest commercial people'. This offers a means by which law and practice might be aligned. But what does 'reasonable expectations of commercial people' mean? In particular, what is the relationship of 'reasonable expectation' and legal doctrine?

II. WHAT IS THE 'LAW OF CONTRACT'?

There are several criteria by which we might seek to define the law of contract. One possibility is to define it by content so that the subject is defined as comprising those topics normally covered in texts and courses on the subject. An alternative would be to define it by subject matter as the law applicable to contracts. This immediately poses the question: 'what is a contract?' If the answer is that a contract is a legally enforceable agreement the definition is immediately open to several objections: first, that it is circular, for by what criteria do we determine that an agreement is legally enforceable; second, that it is too narrow, for the law of contract as normally defined by the contents of texts and courses includes material concerned with the effects of non-binding agreements; and third, that it is, conversely, too wide, for the law applicable to contracts may include topics normally thought of as belonging to other substantive areas of law, such as restitution, tort, property and trust. There is, then, a difference between the law of contract as defined by the texts and the law applicable to contracts. If we define the law of contract as the law applicable to contracts, is the emphasis on the work of lawyers dealing with contracts, or on contracts as social and commercial phenomena? And in either case, is the emphasis on the processes by which contracts come about—transactional law—or on the resolution of disputes arising out of contracts?

It is tempting to concentrate on a content-based definition. In some areas of legal study there is a considerable degree of variance between the contents of different University courses and text books on ostensibly the same subject. For instance although most commercial law courses and texts cover the law of sale(s) they cover a wide range of other subjects, including agency, insurance, competition law, banking and finance, commercial credit and security, consumer credit, international sales, including carriage and finance, and so on. In contrast the contents of contract law courses and, even more so, texts exhibit a large degree of similarity. A glance at the contents pages of the three 'classic' student texts, *Anson*,[1] *Cheshire Fifoot and Furmston*[2] and *Treitel*,[3] illustrates this.[4] The same topics are covered in more or less the same order and with more or less the same proportion of the text devoted to each. All three books start[5] with an examination of '... formation of contracts' (covering offer and acceptance, consideration, intention to create legal relations, certainty and formalities); then contents of the contract (contract terms, express and implied, their classification, and exclusion clauses and controls on their effectiveness); factors affecting validity of contracts (including mistake, misrepresentation, duress and

[1] Beatson, *Anson's Law of Contract*, 28th edn, (2002).
[2] Furmston, *Cheshire, Fifoot and Furmston's Law of Contract*, 14th edn, (2001)
[3] Treitel, *The Law of Contract*, 10th edn, (1999).
[4] The contents of the leading practitioner text, Beale (ed) *Chitty on Contracts* are similar. See also Furmston (ed), *Butterworth's Law of Contract* (1999).
[5] *Anson* and *Cheshire, Fifoot and Furmston* both have a short introductory section dealing with broad themes, principles and historical development.

undue influence, and illegality and public policy); privity of contract and related issues (including agency and assignment); discharge of contracts (covering performance, breach and frustration); and, finally, remedies. The respective styles differ but the general approach is the same, with an emphasis on doctrine, seeking in the classic common law style to extrapolate general rules from decisions in individual cases.

The approach of these 'classic' texts matters because they exert a strong influence on not only academic but also practitioner thinking. Professors, practitioners and judges, we have all been raised on them, and they are used not only as student but also as practitioner texts. The classical structure they adopt is seductive. Material needs to be organised for the purposes of exposition and the classical structure is familiar and, on the whole, it works. Part of its attraction is that it appears to tell a linear story which begins with formation and ends with the culmination of the contract either in performance or in breach and its consequences. But this is by no means the only way the material might be organised. Our approach might differ according to whether our emphasis is on transactions or on dispute resolution. For instance, if the focus is on dispute resolution and the law in cases, the logical place to start might be at the end, with damages. After all, the law of damages for breach of contract, with its central objective of protecting the claimant's expectation, offers a nice introduction to the general objective of the law of contract. Moreover, a contract dispute normally involves a claim for damages for breach, to which arguments of no agreement/consideration/intention to create legal relations, frustration, non-performance justified by breach and so on are put forward as defences. If, on the other hand, our focus is on transactions we might start with negotiation and initial formation, separating out topics such as frustration and implied terms as both being concerned with the effects of events not provided for in the contract, and perhaps gathering them together with others such as promissory estoppel, variation, and waiver in an examination of the ways in which contractors deal with changed circumstances. Exclusion clauses could sensibly be hived off from the section on terms either to the section dealing with performance, as terms defining performance, or to an entirely separate section, whilst the section on classification of terms as conditions, warranties and innominate terms might usefully be joined with the section on discharge by breach, where it has most relevance. These changes would not necessarily produce a better organisational structure but they are genuine alternatives and they are not self-evidently worse than the classical structure.[6]

[6] See Professor Halson's *Contract Law* (2001). Halson divides his book into five main sections: The Negotiation Stage, covering such matters as mistake, misrepresentation, undue influence, proprietary estoppel and so on; The Birth of the Contract, subdivided into three sub-sections, covering Positive Requirements (offer, acceptance, consideration etc), Negative factors (illegality and some forms of mistake) and Personnel (privity and capacity); The Life of the Contract, covering content (terms), modification of the contract (including promissory estoppel, duress

Why is there such a large degree of similarity in approaches to the subject? The similarity in the content of courses can of course be explained in part by the need to satisfy the requirements of the legal professional bodies, but that is only a small part of the explanation. In recent years those requirements have been relaxed so that the content of individual University subject courses is no longer prescribed in any detail.[7] The organisational structure of the law of contract was bequeathed to us by the text book writers of the late 19th century, but since that time there have been innumerable changes—the development of a separate requirement of intention to create legal relations and of the doctrines of promissory estoppel and economic duress, the changed perception of the basis of the doctrine of frustration, the development of controls on exclusion clauses and so on. Moreover there has been a change in the perception of the underlying philosophy of contract, with the development of statutory controls on a whole range of contracts and contract terms calling into question such classical principles as freedom and its corollary, sanctity of contract. The hoary old question 'law of contract or law of contracts?' has a new vitality as the subject may now be stratified in more than one dimension, not only into the law of sale, credit, agency, employment and so on, but into market sectors with divisions between domestic and international contracts, consumer and business contracting and so on.[8]

A. The Law Applicable to Contracts

If contract law is the law concerned with contracts we might expect to see some correlation between its content and the work of practitioners dealing with contracts. In fact, however, the selection of material for study in the classic texts does not reflect the importance of that material in practice. The texts seem to place too much emphasis on some topics of little practical relevance, whilst under-emphasising others and omitting some topics of real practical relevance altogether.

and frustration) and performance; the Death of the Contract (frustration and discharge by breach); and The Aftermath, covering remedies.

[7] See the Joint Statement issued by the Law Society and the General Council of the Bar on the Completion of the Initial Academic Stage of Training. In passing we may question whether the topics prescribed for study are appropriate for students intending to enter legal practice. Interestingly practitioners regularly complain that newly recruited trainees know insufficient contract law, although whether this is simply a reflection of the fact that in most cases the student last studied contract law over three years before entering practice, and forgot that the morning after the examination, or an indication that they studied the wrong material in the first place is another matter.

[8] These developments have led some commentators to talk of the emergence of a 'modern' law of contract whose doctrines, principles and underlying values are distinct from those of the 'classical' law. See, eg, Adams and Brownsword *Understanding Contract Law* 3rd edn, (2000).

Clearly there is a limit to what can be studied in a contract course and the degree to which we can or should expect student texts and courses of study to track legal practice. Contract will normally be taken by students in their first year of legal study, and in Universities where the degree course is modular and semesterised, may be studied for only twelve weeks. Moreover the contract course typically carries a heavy burden, serving not only to introduce students to one of the fundamental areas of legal study but also to introduce them to case law method and statutory interpretation and at the same time to provide a basic understanding of relevant legal doctrine whilst placing the law in its historical, social and economic context. Our objective is not merely to produce doctrinal technicians but to encourage students to think critically about the law. We should not further load our students by expecting them in their first year of study to study the details of particular commercial transactions: the objective of the contract law course is to introduce students to the general principles of the subject, leaving more detailed study of individual contract types to later courses.[9] We must also recognise that academic study inevitably lags behind commercial practice so that new developments will only be considered in text books some time after their emergence in practice and subsequent consideration by the courts.

B. Textbook Law and Practice

We might nevertheless expect to see some correlation between contracts in practice and the contents of text books. Insofar as the text books do reflect the experience of the commercial practitioner they reflect only part of that work. The law in the books is concerned not so much with living contracts but with the pathology of failed contracts.[10] Text books focus on the law as expounded in the higher courts, concerned with that tiny fraction of the multitude of contracts made every day which break down and give rise to litigation which goes to trial. Almost by definition a disproportionate volume of the reported cases will be concerned with high value contracts. In short, insofar as the books are concerned with the work of the practitioner, it is the work of the litigator rather than of the commercial contract drafter. Of course, the drafter needs to know the default rules which will apply in

[9] Perhaps there is a need for a wider range of advanced contract courses, studying aspects of typical commercial transactions, as suggested by Professor Bridge in a paper given at the SPTL conference in Glasgow in September 2001.

[10] Interestingly several of the cases concerned with the most fundamental rules of contract formation are not concerned with contract disputes at all but with the application of criminal, regulatory or fiscal rules: see eg, *Grainger & Son v Gough (Surveyor of Taxes)* [1896] AC 325; *Pharmaceutical Society of GB v Boots Cash Chemists (Southern) Ltd* [1953] 1 QB 401, [1953] 1 All ER 482; *Partridge v Crittenden* [1968] 1 WLR 1204; [1968] 2 All ER 421, *Esso v Commissioners of Customs and Excise* [1976] 1 All ER 117; [1976] 1 WLR 1.

the event of a dispute in order to draft an effective contract, but issues such as consideration and mistake will rarely impinge upon his work. But nor does the treatment of these topics in the text books reflect the work of the litigator. It is for instance well known that contract law in the courts is largely concerned with construction,[11] and yet until recently the subject was afforded barely any coverage in the standard texts. Such discussion of the topic as there was, was to be found in the section of the book concerned with exclusion and similar clauses (as, indeed, was most of the material on incorporation of terms),[12] creating the (wrong) impression that those rules were only applicable to exclusion and similar clauses. At the same time the books give extended coverage to topics such as capacity, mistake, illegality and frustration which rarely, if ever, arise in practice, let alone in the case law. Even formation issues are relatively rare and consideration is hardly ever an issue. It may be objected that construction underpins much of the rest of the law in the books, including mistake, frustration, discharge, classification of terms, exclusion clauses and so on, and that construction has therefore always been studied, albeit obliquely. But this merely emphasises the importance of the subject and the peculiarity of its omission from the programme of study as a topic in its own right. Of course, the situation is now changing. The decisions up to and culminating in *Investors Compensation Scheme Ltd v West Bromwich Building Society*[13] alerted academics to the significance of construction and created interest in its study, but the principles in those cases had their origins in cases decided twenty to thirty years ago.[14]

That the preoccupations of the text book writer are not those of the practitioner is nowhere more evident than in the coverage of those topics which are normally grouped together under the heading 'formation of contracts'—offer, acceptance, consideration, intention, certainty and so on. The treatment of these topics in the books bears little relationship to the process of real life contract formation. The books suggest that contract analysis is almost formulaic: 'offer plus acceptance equals agreement; agreement plus consideration equals contract'. The process of reaching an agreement by offer and acceptance appears to resemble a stately, formal dance with a series of ordered steps—invitation to treat, offer, counter-offer, acceptance.

[11] See the observation of Lord Goff in [1984] LMCLQ 382 at p 385.

[12] The latest 10th edn, (1999) of Professor Treitel's admirable text deals with construction wholly in the context of the discussion of exclusion clauses. Professor Treitel devotes nine pages of his chapter on 'The Contents of the Contract' to express terms, eight of them being devoted to the parol evidence rule and its exceptions. In contrast Professor McKendrick in his *Contract Law* 4th edn, (2000) devotes seven pages of his chapter on 'The sources of contract terms' to interpretation.

[13] [1998] 1 All ER 98.

[14] The so-called 'modern' approach is said to originate in the decisions in *Prenn v Simonds* [1971] 3 All ER 237 and *Schuler v Wickman Machine Tool Sales Ltd* [1974] AC 235, [1973] 2 All ER 39 and especially in the speeches therein of Lords Wilberforce and Reid respectively. See McKendrick, 'The Interpretation of Contracts: Lord Hoffmann's Re-Statement', above, ch 7.

The reality, of course, is often altogether more chaotic, even where lawyers are involved.[15] To the practitioner, contract formation is a matter of drafting and negotiation. Terms are proposed, rejected, revised and resurrected. Drafts are amended and re-amended and passed back and forth. Gradually agreement emerges. The process can of course be analysed in terms of offer, acceptance and intention, but, as Lord Wilberforce recognised in *The Eurymedon*[16] it may often be difficult to fit the facts of a real life situation into the text book slots of offer, acceptance and so on. Even a relatively straightforward everyday transaction such as the contract to ride a bus or a sale in a self-service store is difficult to analyse in terms of offer and acceptance.[17] Consideration is rarely, if ever, an issue at the formation stage.[18]

Part of the explanation for this is of course that, as Lord Wilberforce recognised, the 'rules' on 'formation of contracts' do not so much describe the way in which contracts are made, as provide a scheme of analysis which can be used to determine whether a borderline situation gives rise to a legal obligation enforceable by the law of contract. They are perhaps most useful when the court is asked to infer a contract from conduct. This is not to say that formation cases are rare. Cases are frequently reported which raise formation issues, directly or indirectly. It is just that when these cases arise out of formal negotiations they are rarely concerned with the sort of situations considered in the text books—price lists, advertisements, window displays and so on. Often the cases involve implication of a contract; more often, cases of incomplete or failed negotiations. Sometimes the court is called

[15] See for instance *Carlton Communications plc and Granada Media plc v The Football League* [2002] EWHC 1650 (Comm), discussed below, where negotiations lasted some three months and included exchanges by letter, fax, e-mail, telephone and face to face meeting.

[16] *New Zealand Shipping Co Ltd v A M Satterthwaite & Co Ltd, The Eurymedon* [1975] AC 154 at p 167.

[17] See the comments of Stephen J in *MacRobertson Miller Airline Services v Commissioner of State Taxation of State of Western Australia* [1975] 8 ALR 131, HC of Australia, at p 139.

[18] Significantly the leading modern case on consideration, *Williams v Roffey Bros and Nicholls (Contractors) Ltd* [1991] 1 QB 1, [1990] 1 All ER 512 was concerned not with contract formation but with contract variation, and the argument on consideration was only added at a relatively late stage in the proceedings. Another leading case, *Pao On v Lau Yiu Long* [1980] AC 614, [1979] 3 All ER 65, PC was also a variation case. Yet another, *The Eurymedon*, above n 16, was concerned with an inferred contract. *Williams v Roffey* is a prime example of the different emphases of academics and practitioners. The decision, which modified the old 'rule' in *Stilk v Myrick*, was subject to detailed academic analysis when it was handed down in 1990 and hailed as one of the most significant developments in the law of contract in recent years. However, twelve years on its practical impact has been minimal. A Lexis search for the case in 2002 revealed just twelve citations of the case in the intervening period, and only one— *Simon Container Machinery Ltd v Emba Machinery AB* [1998] 2 Lloyd's Rep 429—directly on the point addressed in the case. Compare this with the impact of the *Investors Compensation Scheme* case as revealed in Professor McKendrick's essay, 'The Interpretation of Contracts: Lord Hoffmann's Re-Statement', above, ch 7, which notes that in the four years since that decision was handed down it has been cited 183 times. The explanation is no doubt partly that *Roffey* merely reflected what was already the normal business practice.

upon to analyse a course of negotiations and determine whether a contract ever came into being. In others it is not disputed that there is no contract, but the court is called upon to provide a remedy under some other heading to a party who has incurred expenditure in anticipation of a contract or otherwise suffered loss due to the breakdown of negotiations.[19] Such cases often involve consideration of the legal effects of documents and procedures not considered in the texts, such as 'heads of agreement' and 'letters of intent', and of legal doctrines from beyond the boundaries of contract law, such as restitution, tort, proprietary estoppel and constructive trust.

Notwithstanding their relative scarcity in the reports, however, the formation rules are important for several reasons. First, they are familiar—so much so that they are often taken for granted and as settled. Secondly, they provide most students with their introduction not just to the law of contract, but to common law in general. Thirdly, although formation disputes are comparatively rare, the moment of formation is critical in relation to many issues. At the moment of contract formation the parties' relationship is crystallised. Issues of interpretation, frustration, implied terms and remoteness of damage depend in one way or another on the intentions of the parties at that moment. Moreover, it seems that at the moment of formation the parties' relationship changes. Until that moment parties remain largely free, subject only to the limitations imposed by the rules on misrepresentation, duress, undue influence and so on, to pursue their own interests in an aggressively individualistic fashion.[20] Thereafter, bound together by contract, their freedom may to some degree be circumscribed by implied duties including the duty of cooperation.[21] Fourth, the same rules are, on the whole, applied to contract renegotiation, even though renegotiation between parties already bound by a contract raises issues very different from those raised by initial negotiation.

We might therefore criticise the text book analysis of contract formation on several grounds. It focuses on issues of no practical importance, whilst omitting others of real commercial significance. It runs together issues of contract formation and contract renegotiation, despite the fact that cases of initial formation and renegotiation raise very different policy considerations.

[19] Such cases are, apparently, not uncommon, especially in the context of building and engineering contracts. See the opening remarks of His Honour Judge Hicks QC in *Hescorp Italia SpA v Morrison Construction Ltd* (2000) 75 Con LR 51, a case of failed negotiations: 'The background to the issue which I have to decide is one all too familiar in this court.'

[20] *Walford v Miles* [1992] AC 128, [1992] 1 All ER 453; *CTN Cash and Carry v Gallagher* [1994] 4 All ER 714. There are of course restrictions on the freedom of action of negotiating parties, imposed by doctrines of duress, misrepresentation and so on, which may be seen as aspects of a duty of good faith. See also the comment of Brooke LJ in *Laceys Footwear Ltd v Bowler International Freight Ltd* [1997] 2 Lloyd's LR 367 at p 385.

[21] Whilst duties of good faith may not be imposed on negotiating parties there are signs of a growing willingness to recognise such duties—even if not expressly named as such—on parties to a contract. See *Phillips Electronique Grand Public SA v British Sky Broadcasting Ltd* [1995] EMLR 472; *Paragon Finance Ltd v Nash*[2001] EWCA 1466; [2001] 2 All ER (Comm) 1025.

Above all, however, we might say that, all too often, the formation 'rules' are overstated, presented as abstract rules disconnected from any underlying principle or rationale, capable of mechanistic application without thought. Thus we are told that 'a promise is not binding unless supported by consideration'; that 'an agreement to agree is not binding because it is incomplete and uncertain'; that 'a shop window display or advertisement of goods for sale is not an offer'; 'silence cannot amount to acceptance'; 'a posted acceptance is effective on posting'; 'a person is bound by terms in a document they sign'; 'estoppel may operate as a shield but not a sword'; and so on. Most of these statements, if expressed in these bald terms, are misleading, if not plain wrong.

Does this matter? The leading texts all make clear that these 'rules' are no more than presumptions, prefixed by the word 'generally' and hedged about with qualifications. The answer is that it does, nevertheless, matter, for a number of reasons. We must first consider the 'trickle down' effect. Rules of law are used not merely to determine the outcomes of cases before the courts, but to provide a framework for resolution of disputes without resort to litigation and, more broadly, to guide the behaviour of contractors who, aware of the default rules, can structure their contracts and procedures accordingly. If the rule is known, the outcome of litigation can be predicted; but the outcome can only be predicted accurately if the rules are stated accurately. Contract law is business law. When the rules reach the community of business people and non-legal professionals they tend to do so in absolute form, stripped of subtlety and nuance. There is therefore a tendency for some, including some who should know better, to treat some of these 'rules' as absolute. Take, for example, the 'rule' in *L'Estrange v Graucob*[22] that a person who signs a document is bound by its contents. The underlying rationale, surely, is that, objectively viewed, a person who signs a document thereby makes it appear to the other contracting party that he assents to the terms it contains, and yet it is often treated as an absolute rule: 'get your terms signed and they apply'. Thus when the Court of Appeal held in *Harvey v Ventilatoren Fabrik Oelde GmbH*[23] that a signatory was not bound by terms in a signed document presented for signature which was misleading as to its contents, there was considerable surprise, and yet this is entirely consistent with the suggested underlying rationale.

An understanding of the principle underlying a rule is also needed when it becomes necessary to apply the rule to new situations. Much has been written in recent years about the application of contract formation rules to electronic commerce. It has widely been assumed that an offer of goods or services for sale on a web site is not a contractual offer on the analogy with

[22] [1934] 2 KB 394.
[23] [1988] BTLR 138. See also *Grogan v Robin Meredith Plant Hire Ltd* [1996] CLC 1127. *Harvey* is not cited by any of the 'classic' texts.

the supposed rule for shop displays.[24] But is this necessarily the case? The established 'rule' is based on the presumed intention of the parties, influenced in part by recognition that if an advertisement is an offer the supplier may have problems in supplying enough goods to meet demand. But what if the item 'offered' on a web site is intangible, such as downloadable software? Why should the web site then not be considered to make an offer? Why indeed is it assumed that a traditional window display is not an offer? It is clear that in appropriate circumstances it may be.[25]

Treatment of the rules as absolutes also causes misunderstanding when we come to compare our laws with those of other systems. To take but one example, it is sometimes suggested that the rules of the UN Vienna Convention on Contracts for the International Sale of Goods on contract formation are different from those of English law. The Convention, for instance, does not follow the English postal rule but provides that a posted acceptance is not effective until it reaches the offeror and allows an acceptance to be revoked if the revocation reaches the offeror before the acceptance.[26] However, on closer examination the differences are more imagined than real. The Convention also states that an offer can be revoked until it is accepted but that the revocation must reach the offeree before acceptance is dispatched, thus replicating in the most important situation the effect of the postal rule.[27] And as for the revocability of a posted acceptance, there is in fact no English authority on the point. The presumption against revocability is based simply on extrapolation from the so-called postal 'rule', a rule originally devised to determine priority as between a posted acceptance and a posted revocation of an offer. There is—as in fairness the text books recognise—no good reason to hold that a posted acceptance can never be revoked.

III. THE LAW AND BUSINESS PRACTICE

There is then sometimes a mismatch between the text book 'rules' of contract law and the legal practitioner's experience. Of rather more concern is that there is sometimes a second mismatch between the law itself and business practice. Examples are frequently thrown up by the case law. In the context of the rules governing contract formation, examples are provided by the

[24] *Fisher v Bell* [1961] 1 QB 394.
[25] Cf *the Principles of European Contract Law (PECL)* Art 2.201 under which a proposal to supply goods or services made by a professional by a display of goods is presumed to be an offer to sell or supply at that price until the stock of goods, or the supplier's capacity to supply the service, is exhausted.
[26] *CISG* Art 22; *UNIDROIT Principles of International Commercial Contracts* Art 2.10; *PECL* Art 1.303; and see note to Art 2.205.
[27] *CISG* Art 16(1); *UNIDROIT Principles* Art 2.4; *PECL* Art 2.202.

cases concerned with certainty and completeness where it is argued that an agreement, often one acted on and treated as enforceable for a period of time, is in reality not a contract at all because it lacks the requisite degree of certainty. Similarly the battle of the forms case law demonstrates the difficulties in applying traditional 'offer/counter-offer/acceptance' analysis to real life negotiations.

The mismatch between the text book rules on formation and business practice is, of course, well known. Beale and Dugdale noted in 1975 that the contracting practices of some businesses examined in their survey would often not result in the creation of an effective contract and observed that:

> there was considerable awareness of the fact that in many cases an exchange of conditions would not necessarily lead to an enforceable contract, and in some that the last set of conditions might prevail ... But most firms seemed unconcerned about the failure to make a contract ... Legal enforceability seemed secondary to reaching a common understanding.[28]

This mismatch is recognised by the courts[29] which seek, by manipulation of doctrine, to minimise its effects. Thus the doctrine of consideration is capable of almost limitless manipulation to allow consideration to be found, or not found, as the justice of the case requires.[30] A lack of certainty in an agreement will not prevent its being upheld as a contract, especially where the parties have acted on the basis that they have a binding agreement. The courts recognise that:

> When two businessmen wish to conclude a bargain but find that on some particular aspect of it they cannot agree ... it is not uncommon for them to adopt language of deliberate equivocation, so that the contract may be signed and their main objective achieved. No doubt they console themselves with the thought that all will go well and that the term in question will never come into operation or encounter scrutiny.[31]

The preferred approach is that 'the dealings of men may so far as possible be treated as effective, and that the law may not incur the reproach of being

[28] 'Contracts between businessmen: planning and the use of contractual remedies' (1975) 2 British Journal of Law and Society 45 at p 50.

[29] See for example Lord Wilberforce in *New Zealand Shipping Co Ltd v Satterthwaite & Co Ltd* [1975] AC 154 at 167, PC. See also *Pagnan SpA v Feed Products Ltd* [1987] 2 Lloyd's Rep 601 at 610–611 (Bingham J), 618–619 (Lloyd LJ); *MacRobertson Miller Airline Services v Commissioner of State Taxation of State of Western Australia* [1975] 8 ALR 131, HC of Australia per Stephen J at 139.

[30] It was of course that flexibility which allowed the Court of Appeal in *Williams v Roffey Bros and Nicholls (Contractors) Ltd* to realign the doctrine of consideration with business practice.

[31] Staughton J in *Chemco Leasing SpA v Redifusion plc* (1985) unreported, *Lexis* transcript, quoted by Hirst J in *Kleinwort Benson Ltd v Malaysia Mining Corpn* [1988] 1 All ER 714 at 720.

the destroyer of bargains'.[32] Even a failure to agree important terms may not be fatal: the court has extensive power to fill in gaps in the agreed terms by implying the terms necessary to make the contract work.[33]

Does it then matter that there is a mismatch between text book law and business practice if the courts are generally prepared to recognise that fact and apply the text book rules in a flexible manner? It is submitted that it does, for a number of reasons. First, this approach tends to compound the problem referred to earlier, that the description of the rules in the books does not accurately reflect the reality of the rule in application. If the court manipulates doctrine to fit the facts of the particular case it becomes more difficult to use precedent to predict the outcome of future disputes, unless the reasons for the manipulation are identified and articulated. Second, the mismatch may encourage the deployment of technical legal arguments lacking in commercial merit. Consider, for instance, the line of cases such as *Hillas v Arcos*[34] and *Foley v Classique Coaches Ltd*.[35] In these cases parties enter into an agreement for the supply of goods or services over a period of time. Clearly the parties initially think they have a 'deal', and for a time behave accordingly. Only later when for one reason or another the original deal becomes less attractive is it argued that the deal was uncertain and therefore not a binding contract. In such circumstances the uncertainty argument is a lawyer's argument, posited no doubt when one party defaults on the deal and is sued by the other for breach, and clearly lacks merit. It is not surprising to find the courts striving to find ways to uphold the contract and cure any uncertainty. A striking case is *G Percy Trentham Ltd v Archital Luxfer Ltd*.[36] In this case T were contractors engaged on a building contract. They entered into negotiations with A for A to act as sub-contractors on the project and asked A to commence work in anticipation of completion of the contract. Negotiations continued over a three month period, with some issues never being expressly resolved. Nevertheless the sub-contract work was completed and paid for.

[32] Per Lord Tomlin in *Hillas v Arcos* [1932] All ER Rep 494 at 499.

[33] There are limits to this approach. Its application depends on the court finding that the parties have entered into an agreement with the objectively determined intention that agreement should be binding. The very failure to agree crucial terms is evidence that they do not have the requisite intention: see Lloyd LJ in *Pagnan SpA v Feed Products Ltd* [1987] 2 Lloyd's Rep 601 at 619. The reasoning process is therefore essentially circular. Inevitably, therefore, there is an element of uncertainty engendered by the very flexibility of this approach. Moreover, it breaks down where the parties acknowledge that issues remain to be agreed between them, as in *May and Butcher v R* [1934] 2 KB 17. This is exemplified in the refusal in *Walford v Miles* [1992] AC 128, [1992] 1 All ER 453 to recognise an agreement to negotiate in good faith, an issue to which I shall return below. This issue is at the core of the decision in *Baird Textile Holdings Ltd v Marks & Spencer plc* [2001] EWCA Civ 274, [2002] 2 All ER (Comm) 193, discussed below.

[34] [1932] All ER Rep 494.

[35] [1934] 2 KB 1. For a more recent example see *Mamidoil-Jetoil Greek Petroleum Co SA v Okta Crude Oil Refinery AD* [2001] EWCA Civ 406, [2001] 2 All ER (Comm) 193.

[36] [1993] 1 Lloyd's Rep 25.

Subsequently, however, the employer brought a claim for damages for breach of contract against T and T sought to claim an indemnity from A, alleging that there were defects in the work carried out by A. A now claimed as part of their defence that there was no contract between the parties on the grounds that the parties had never reached final and complete agreement. In the circumstances this argument had, as the Court of Appeal recognised, an air of unreality. The facts of the case resemble in some ways those of *British Steel Corpn v Cleveland Bridge and Engineering Co Ltd*[37] in which performance began pursuant to a letter of intent before all the terms of the contract had been agreed and Robert Goff J held that there was no concluded contract between the parties. In *Trentham* however, the key aspects of the parties' transaction had been fully performed on both sides, and the Court of Appeal held that a contract had come into being at the latest during performance, even if it was not possible to analyse it precisely in terms of offer and acceptance or to identify the precise moment at which it came into being.

Such situations are probably not uncommon. The Court of Appeal no doubt reached the right result on the facts. The point for present purposes is that again the text book rules encouraged the deployment of a highly technical legal argument with no commercial merit.

Williams v Roffey provides a further example. The facts are well known and, again, probably not untypical. The case arose out of renegotiation of a building sub-contract. The defendant having agreed to increase the sums payable to the sub-contractor defaulted on the revised agreement and the sub-contractor sought to enforce it. Various defences were put forward to resist the claim. The defence of 'no consideration' was added, more or less as an afterthought, not having been pleaded in the original defence. 'No consideration' is the argument of the lawyer, not of the builder. Clearly, in the language of business, the parties thought they had 'a deal'. In a commercial context arguments of 'no consideration' rarely have any real merit, especially now that the risks of abusive renegotiation may be guarded against by the doctrine of economic duress, and may be quickly disposed of. Indeed, post *Roffey* it will generally be relatively easy to 'find' consideration to support a promise where a court wishes to do so.

Cases such as *Foley*, *Roffey* and *Trentham* illustrate the difficulty of applying the text book rules of contract formation to real life contract situations. It may be objected that the arguments deployed in *Foley*, *Roffey* and *Trentham* failed, demonstrating that flexible application of the formation rules can minimise the effects of the mismatch between law and business practice. The point, however, is that those arguments were deployed in the first place. Similar arguments are no doubt deployed in cases which do not reach court.

[37] [1984] 1 All ER 504.

In the same way the mismatch between law and business practice encourages the use of 'dodges, tricks and ruses' in the process of contract formation itself. This may be seen in the tactics used to try and win the so-called 'battle of forms'. In *Butler Machine Tool Co Ltd v Ex-Cell-O Corpn*[38] the battle was lost by the sellers returning the tear-off slip from the buyer's order form. It is most unlikely that the sellers intended by so doing to accept the buyer's terms. The rationale for the decision is supposedly that *viewed objectively* the seller's conduct indicated an intention to assent to the buyer's terms. But in point of fact the seller's conduct in *Butler* was at best equivocal, for the slip was returned with a letter referring to the sellers' own terms, and a reasonable person in the buyer's position would almost certainly realise that it was at least as likely that the slip was returned inadvertently, by a relatively junior employee, with no intention or authority to waive the seller's terms.[39] The use of the slip in *Butler* may not have been intended to trick the seller into agreeing to the buyer's terms, but the use of such slips in that way borders on sharp practice. The same can be said of other tactics used to try and win the battle. The battle is won by the side which fires the last shot. The seller may seek to take advantage of this rule by delivering the goods with a delivery note including its terms, arguing that by receiving the goods the buyer has accepted its terms. In order to avoid losing the battle in this way, buyers are advised to stamp delivery notes with an indication that the goods are received 'subject to our terms'. A more sophisticated tactic is the use of a clause paramount in the standard terms by which the business indicates that it will contract only on its own terms and will not agree to any variation of them. The argument goes that, having indicated in its standard terms that it will not agree to any variation in those terms the business cannot be taken to have agreed to any different terms. But in everyday business life terms in standard form documents will often go unread.[40] Again the justification for the approach is the objective approach to consent. But if experience shows that businesses tend not to read standard term documents, how can it be said that it is reasonable to expect them to do so? The objective reality in many cases is surely that neither party agrees to the other's terms, but a conclusion that the parties have made no contract in a case such as *Butler* is as unrealistic as it would be on the facts of *Trentham*.[41] The objection is to the rigid application of offer-acceptance analysis to reach an equally unrealistic conclusion that one party has agreed to the other's terms. The only sensible solution applying

[38] [1979] 1 All ER 965, [1979] 1 WLR 401, CA; see also *British Road Services Ltd v Arthur Crutchley & Co Ltd* [1968] 1 All ER 811 and *Sauter Automation v Goodman* (1986) 34 BLR 81.

[39] See to this effect *Commissioners of Inland Revenue v Fry* [2001] STC 1715.

[40] See Beale and Dugdale *op cit* n 33.

[41] Steyn LJ in *Trentham* noted that the facts of the case shared some features with the battle of forms cases.

traditional analysis is that the parties have agreed a contract without either agreeing to the other's terms, so that there is, in effect, an open contract.[42] And yet that too is clearly inconsistent with the parties' intentions. Perhaps there is something to be said here for the approach suggested by Lord Denning in *Butler*, requiring the court to construct a contract from the two parties' terms.[43]

Essentially the criticism of the mismatch between text book law and business practice is thus that it encourages opportunistic behaviour. In the context of the formation rules that tendency may manifest itself in the form of an attempt to take advantage of the formation rules to secure an unintended agreement, or an ex post facto attempt to use the rules to escape from what was thought to be a binding contract. Indeed, the House of Lords seems to have sanctioned opportunistic behaviour between negotiating partners by its refusal in *Walford v Miles*[44] to enforce an agreement to negotiate in good faith. *Walford* has of course been criticised, in particular by commentators who advocate the adoption of a general principle of good faith in English contract law.[45] If the courts are willing to recognise an obligation not to exercise a contractual discretion 'arbitrarily, capriciously or for an improper purpose'[46] why can they not recognise an obligation not to break off negotiations arbitrarily, capriciously or for an improper purpose? One response might be that once a contract is concluded the parties to it are engaged in a co-operative venture whereas at the negotiation stage the parties are in an adversarial situation in which each 'is entitled to pursue his or her own interest'. But this is not entirely satisfying.[47] First it is to

[42] See *Lidl UK GmbH v Hertford Foods Ltd* [2001] EWCA Civ 938.
[43] See [1979] 1 All ER 965 at 968. Cf *the UNIDROIT Principles* Art 2.22 and *PECL* 2.209 which provide, in effect, that where parties reach agreement but refer to different standard terms, a contract is formed which incorporates those terms which are 'common in substance' unless either party indicates in advance, or without delay (undue delay—*PECL*) informs the other that it does not intend to be bound by the contract.
[44] [1992] AC 128, [1992] 1 All ER 453.
[45] Lord Steyn has suggested that the decision might be reconsidered: (1997) 113 LQR 433 at 439. Compare the UNIDROIT Principles of International Commercial Contracts Art 2.15 and the Principles of European Contract Law Art 2.301, both of which impose liability on a party who breaks off negotiations in bad faith. *Walford* creates difficulties beyond the stage of initial formation. So long as English law does not recognise an agreement to negotiate, it is incapable of enforcing so-called hardship clauses often found in commercial contracts: see McKendrick in Forte (ed) *Good Faith in Contract and Property Law*, (Oxford, Hart, 1999). On hardship clauses generally see Schmitthoff [1980] JBL 82; and see Hooley, 'Material Adverse Change Clauses After 9/11', above, ch 12. Similarly it has been held that a dispute resolution procedure requiring the parties first to attempt to resolve disputes by negotiation is unenforceable on the basis of *Walford v Miles:* see *Halifax Financial Services Ltd v Intuitive Systems Ltd* [1999] 1 All ER Comm 303. But cf *Cable & Wireless plc v IBM United Kingdom Ltd* [2002] EWHC 2059 (Comm), [2002] 2 All ER (Comm) 1041.
[46] *Paragon Finance Ltd v Nash* [2001] EWCA 1466; [2001] 2 All ER (Comm) 1025.
[47] There is some support for such a distinction in the judgment of Sir Thomas Bingham MR in *Phillips Electronique Grand Public SA v British Sky Broadcasting Ltd* [1995] EMLR 472 where he indicated that he would have been prepared to imply a term requiring contracting parties to act in good faith towards each other in performance of the contract.

some degree circular. The vehicle for recognition of a duty to negotiate would presumably be an implied contract so the rationalisation becomes 'no contract to negotiate can be recognised because there is no contract yet'. In any case, why should the moment of formation be so crucial when in many cases, that moment is so difficult to identify? How could such a distinction apply in a case such as *Trentham* where performance begins before negotiations are concluded and a contract comes into being during the course of performance? Could it then be argued that the parties are subject to an implied duty to negotiate in good faith to resolve the outstanding matters?

IV. UPHOLDING REASONABLE EXPECTATIONS

Trentham v Archital Luxfer[48] provides a significant example of the courts' willingness to adopt a flexible approach to the application of the contract formation rules in practice. In that case Steyn LJ, as he then was, giving the judgment of the court, explained that 'the governing criterion [in relation to the issue of contract formation] is the reasonable expectations of honest men.' He repeated the same idea, in more detail in *First Energy (UK) Ltd v Hungarian International Bank Ltd*[49] when he opened his judgment as follows:

> A theme that runs through our law of contract is that the reasonable expectations of honest men must be protected. It is not a rule or principle of law. It is the objective which has been and still is the principal moulding force of our law of contract. It affords no licence to a judge to depart from binding precedent. On the other hand, if the *prima facie* solution to a problem runs counter to the reasonable expectations of honest men, this criterion sometimes requires a rigorous re-examination of the problem to ascertain whether the law does indeed compel demonstrable unfairness.[50]

The idea that the purpose of contract law is the protection and promotion of reasonable expectations is not new.[51] It is however rare to find it articulated explicitly by a senior member of the judiciary. His Lordship has more recently explored this notion at more length in an extra judicial context, where he said 'a thread runs through our contract law that effect must be given to the reasonable expectations of honest men.'[52] Similar ideas can be found, albeit less explicitly, in pronouncements by other senior members of

[48] [1993] 1 Lloyd's Rep 25.
[49] [1993] 2 Lloyd's Rep 194. See also *Associated Japanese Bank (International) Ltd v Credit du Nord SA* [1988] 3 All ER 902 at 903.
[50] [1993] 2 Lloyd's Rep 194 at 196.
[51] See especially Reiter and Swan, 'Contracts and the Protection of Reasonable Expectations' in Reiter and Swan (eds) *Studies in Contract*, (1980).
[52] (1997) 113 LQR 433.

the judiciary. Most notable, perhaps, is the judgment of Sir Thomas Bingham as Master of the Rolls in the *Blackpool Aerodrome*[53] case in which the Court of Appeal held that a local authority which invited tenders for the operation of an aerodrome and specified the tendering procedure to be followed by bidders, was contractually bound to consider all conforming tenders delivered by the advertised deadline, on the basis that the invitation for tenders contained an implied offer to that effect, accepted by the plaintiff submitting a conforming tender. Sir Thomas rejected the defendant's contention that it had no legal obligation to consider a conforming tender, observing that if it were accepted 'there would be an unacceptable discrepancy between the law of contract and the confident assumptions of commercial parties'. The tenderer was therefore entitled 'not as a matter of mere expectation but of contractual right'[54] to have a conforming tender considered. Sir Thomas added 'Had the [plaintiff] inquired of the defendant before tendering whether it could rely on a timely and conforming tender being considered ... I feel quite sure that the answer would have been '"of course"'. The law would, I think, be defective if it did not give effect to that.'

A. What is the Status of the Common Theme?

To say that contract law is concerned with upholding and protecting reasonable expectations seems, on one level, to be a truism. Obviously the central function of contract law is the protection of expectations derived from contract. We know, however, that not all expectations are protected. Contract law is concerned with the intentions of contracting parties but intentions are determined objectively, and the touchstone of objectivity is reasonableness. One value of recognising that the objective of contract law is the protection of reasonable expectations is that, at the least, it identifies the principle which underlies the detailed doctrinal rules of contract. The passage from Lord Steyn's judgment in *First Energy*, quoted above, suggests, however that it can provide a means by which to re-align doctrine and business practice. If application of a rule produces a result inconsistent with the reasonable expectations of honest commercial people, then, in the absence of some overriding policy requiring such a result, either the rule or its application must be wrong.

Lord Steyn himself is at pains to deny for his 'common theme' (or thread) the status of a rule or principle of law. It affords no licence to a judge to depart from established precedent. It therefore does not trump established doctrine but provides a litmus test by which to judge the outcomes

[53] *Blackpool & Fylde Aero Club v Blackpool B C* [1990] 3 All ER 25; see Adams and Brownsword (1991) 54 MLR 28.
[54] [1990] 3 All ER 25 at p 30.

of individual cases. This is hardly surprising. English law in general, and the law of contract in particular, has an aversion to broad general principles, preferring instead to rely on specific doctrines offering 'piecemeal solutions' to particular problems.[55] The reason seems to be a fear that acceptance of broad principles would create uncertainty, or (which amounts to much the same thing) that such broad principles could be dangerous in the hands of a maverick judge. Instead the preference of English law is for incremental development of established rules by analogy on a case by case basis. This approach, it is argued, offers greater doctrinal certainty. Broad principles are therefore particularly unattractive in commercial law where it is taken as a given that the law's role is to fulfil the needs of the commercial community and that that objective is jeopardised if the law is uncertain. The certainty of English law, so it is said, is one of the features which make it so attractive to the international trading community, bringing valuable legal business (and other business in its wake) into the English courts.

We might question several of these assumptions. Is it really the case that it is certainty which attracts international legal business, rather than (say) the reputation of English lawyers, arbitrators and judges (or, for that matter, the attractions of London)? Predictability and certainty have been considerably eroded in a number of areas of contract law, with no apparent adverse effect on the popularity of English law.[56] Indeed, we may question whether English law is so certain, fixed and predictable as is sometimes claimed. Recent years have seen a gradual shift from fixed, bright line rules to more flexible doctrines. Witness for instance the shift in the context of contract variation effected by the greater willingness to find consideration for a variation combined with the increased emphasis on the doctrine of economic duress.[57] The rule that performance of an existing duty could be no consideration for a variation promise was a simple, if sometimes hard,

[55] See McKendrick 'The Regulation of Long Term Contracts in English Law' in Forte (ed) *Good Faith in Contract and Property Law*. This aversion is well illustrated by the House of Lords' rejection in *National Westminster Bank plc v Morgan* [1985] AC 686 of Lord Denning's proposal in *Lloyd's Bank Ltd v Bundy* [1975] QB 326 of a power for the court to intervene in contracts on the basis of a general principle of inequality of bargaining power.

[56] For instance, over the last 40 years, since the decision in *Hong Kong Fir Shipping Co Ltd v Kawasaki Kisen Kaisha Ltd* [1962] 2 QB 26, certainty has been considerably eroded in the context of the right to withdraw from a contract on the grounds of breach. Calling a term a 'condition' can no longer be guaranteed to give a right to terminate in the event of its breach: *Schuler AG v Wickman Machine Tool Ltd* [1974] AC 325. Even an express statement to the effect that the contractor may terminate the contract in the event of breach of a particular term may be limited by a finding that the contract should be interpreted as meaning that there will be a right to terminate only for a substantial or serious breach: *Rice v Great Yarmouth B C* (2000) *The Times* 26 July. In the law of sale the buyer's right to reject goods for breach of statutory implied condition has, since 1994, been subject to a restriction where the breach is so slight that rejection would be unreasonable: Sale of Goods Act 1979, s 15A, as amended by Sale and Supply of Goods Act 1994.

[57] *North Ocean Shipping v Hyundai Construction Co, The Atlantic Baron* [1979] QB 705; *Pao On v Yau Liu Long* [1980] AC 614; *Williams v Roffey Bros & Nicholls* [1990] 1 All ER 512.

rule of law. The new rule that performance of an existing duty can be consideration provided that the variation promise is not obtained by duress requires examination of the facts in every case.[58]

The law can then tolerate a measure of uncertainty, even in commercial matters. In any case, one can have too much of a good thing. Absolute doctrinal certainty would make the law rigid and inflexible, incapable of responding to new developments in commercial practice or the facts of particular cases. It is recognised that certainty must be balanced with sufficient flexibility to produce, so far as possible, fair results in individual cases and to respond to new developments, and that flexibility must be bought at the price of certainty.[59] That flexibility is achieved by a willingness to use established doctrines and legal concepts as 'tools' to achieve desired objectives. But using a tool for an unintended purpose may damage the tool or worse. If I use a screwdriver to open a paint tin I may damage the screwdriver (and myself) and in the same way pushing doctrines into new areas may result in their distortion, providing a new source of uncertainty. Moreover, if doctrines are to retain sufficient flexibility to be used as 'tools' they must to some degree themselves be uncertain. The law must be capable of development. On the whole it is at the fringes of established doctrine that the common law develops in its incremental way. We may rightly object to doctrinal uncertainty in cases of conflicting authority, and maybe too in cases where uncertainty is created by a deliberately flexible rule. But we cannot avoid uncertainty at the fringes of doctrine without stultifying the law's ability to develop. And greater unpredictability is created if doctrine rigidly asserted is in fact not applied rigidly. Is a principle of respect for 'reasonable expectations' *necessarily* any more uncertain than the 'piecemeal doctrines', such as the implication of terms, promissory estoppel, the red hand rule of incorporation and so on, by which such expectations are given effect? Recognition of the general principle underlying particular doctrinal rules may provide a guide, for contractors, their advisers and judges, to their future development.

It is submitted that it is a mistake to see broad principles and incremental development as alternatives or rivals. The two can co-exist. The values of a broad principle are that it provides a guide for the application of existing rules, a gap filling mechanism and, by identifying the principle underlying a particular rule or set of rules, a guide to their development. A broad principle of respect for reasonable commercial expectations could thus provide a test by which to judge whether the outcome of particular cases is 'correct'. Where an

[58] A further example is provided by the rule governing the effectiveness of an acceptance sent by 'instantaneous' communication, according to which determination of the time when acceptance becomes effective requires consideration of all the circumstances of the case: see *Brinkibon Ltd v Stahag Stahl GmbH* [1983] 2 AC 34 per Lord Wilberforce at 42.

[59] See Goode, 'The Codification of Commercial Law' (1988) 14 *Monash Law Review* 135 at 150; *Commercial Law in the Next Millennium* p 14ff.

outcome appears not to accord with reasonable commercial expectations the judge is invited to reconsider whether the outcome is dictated by established doctrine, or whether there is scope, by distinguishing an apparent precedent, to produce a result more in accordance with commercial expectation. Where precedent leaves no room for manoeuvre we must conclude that it is the law itself which is out of step with commercial expectation. Presumably then, although Lord Steyn does not expressly say so, the rule should be reconsidered within the bounds permitted by the doctrine of precedent.[60]

B. What are 'Reasonable Expectations'?

What, though, do we mean by 'the reasonable expectations of (honest) commercial people'.[61] Various interpretations are possible. The formulation in *Trentham* would suggest that the outcome of the case should accord with the reasonable expectations of commercial people, implying that the perspective is that of a detached observer. However, the more detailed exposition in *First Energy*, with its reference to reasonable expectations being 'protected' suggests that his lordship had in mind the expectations of one of the contracting parties. This accords with the actual decision in *First Energy* which concerned the offer of a financing facility by the defendant bank to the plaintiff company. The bank argued that there was no contract between the parties on two grounds, first that a facility letter written by the manager of the defendant's Manchester branch could not be construed as an offer and secondly that the manager had no actual or apparent authority to make such an offer. The Court of Appeal decided both issues against the Bank. The facility letter was read as it would be read by 'a reasonable businessman, placed in the same position as' the plaintiff's manager, thus adopting the classic objective approach to interpretation of contractual statements. This could support either interpretation.[62] More guidance is provided by the decision on the authority point. The Court found that the manager had no actual authority to make an offer and was bound by the decision of the House of Lords in *Armagas v Mundogas*[63] to find that he could not by his

[60] The limiting effect of precedent in this context is illustrated by the decisions in *Williams v Roffey* and *re Selectmove Ltd* [1995] 1 WLR 474. The rule in *Foakes v Beer* (1884) 9 App Cas 605 is as inconsistent with reasonable commercial expectation as was that in *Stick v Myrick*, as Lord Blackburn recognised in *Foakes* itself. However, the Court of Appeal in *re Selectmove*, constrained by precedent, was bound to follow *Foakes v Beer* rather than extending the *Roffey* 'practical benefit' concept.

[61] His Lordship has observed that the requirement of honesty adds little to the formulation: (1997) 113 LQR 433 at 434.

[62] Although there is general agreement that contractual intentions are to be determined objectively it is not clear whether this requires a statement to be interpreted from the standpoint of a detached observer or from the standpoint of a reasonable person in the position of the person to whom it was made. It is submitted that in general the latter view is to be preferred. See Furmston (ed), *Butterworths Law of Contract*, paras 2.142–2.147.

[63] *Armagas Ltd v Mundogas SA, The Ocean Frost* [1986] AC 717.

own claim to be authorised give himself apparent authority to make an offer on the Bank's behalf. However, it held that by virtue of his position he had apparent authority to communicate that the bank's head office had approved the grant of the facility sought. Any other decision would 'defeat the reasonable expectations of the parties. And it would fly in the face of the way in which in practice negotiations are conducted between trading banks and trading customers who seek commercial loans.'[64] The rule is thus made to accord with general commercial practice and the outcome of the case to accord with the expectations of a reasonable person in the position of the plaintiffs.

In his *Law Quarterly Review* article Lord Steyn expounds the meaning of 'reasonable expectations' in more detail. Citing Reiter and Swan[65] he explains that:

> The expectations which will be protected are those that are, in an objective sense, common to both parties. The law of contract is generally not concerned with the subjective expectations of a party. The law does not protect unreasonable expectations. It protects only expectations which satisfy an objective criterion of reasonableness. Reasonableness is a familiar concept and no definition is necessary. [It] postulates community values. It refers not to the standards of Lord Eldon's day. It is concerned with contemporary standards not of moral philosophers but of ordinary right thinking people. Sometimes those standards will receive their distinctive colour from the context of a consumer transaction, a business transaction, or even a transnational financial transaction. And the usages and practices of dealings in those disparate fields will be prime evidence of what is reasonable.[66]

In short, when both parties share the same expectations there is no difficulty. In the unlikely event of a dispute arising the law upholds their shared expectation. In the problem case where the parties do not share the same expectation, A will be required to respect B's expectation if either he knows of it or if, in all the circumstances, he ought, as a reasonable person to know of it. And if B's expectation is one a reasonable person would share, A, as a reasonable person, ought to be aware of it.[67]

C. Reasonable Expectations and Legal Doctrine

Many of the rules in the contract rule book can be seen as applications of the general principle of respect for reasonable expectations. The objective

[64] At p 204.
[65] *Op cit* above n 51.
[66] (1997) 113 LQR 433 at 434.
[67] Expressed in these terms there is little difference between a principle of upholding reasonable expectations and a requirement of good faith: if A knows, or ought reasonably to know of B's expectations, A cannot in good faith act otherwise than in accordance with them.

approach to the ascertainment of contractual intention and interpretation of contracts is an obvious example. So too are the rules governing the implication of terms. The 'red hand' rule on incorporation of unusual terms is another example, and we could go on. But it is clear from Lord Steyn's exposition of the 'reasonable expectations' principle that he sees it as more than simply describing the effect of existing rules. As his judgment in *First Energy* shows, a principle of upholding reasonable commercial expectations can be used to realign the outcomes of cases, and even contract doctrine, with commercial practice. Where the application of an established rule produces a result which differs from reasonable commercial expectation we should look again at the rule and its application to decide if the result is compelled by the rule. If practice differs from the rule we should at the very least question the rule. In many cases there will be sufficient elasticity in the rule to allow its application in a way which is consistent with reasonable expectations. Cases such as *Foley*, *Roffey* and *Trentham*, discussed above are examples of this approach. If that is not possible we must be prepared, in the absence of overriding policy considerations, and within the bounds of the doctrine of precedent, to modify the rule book to bring it into line with commercial expectation. If the reasonableness of expectations is to be judged by contemporary standards we must accept that doctrines must be adapted to take account of changes in expectation.

But can a party *reasonably* expect an outcome other than one in accordance with the established contract rule book? Consider here the facts of *Williams v Roffey*. The main contractor's defence to the sub-contractor's claim was that there was no consideration for the promised extra payment. All that the sub-contractor had agreed to do was to perform its existing contractual duties and, according to established precedent, performance of an existing contractual duty owed to the promisor could not be consideration for a fresh promise. In other words any expectation that the sub-contractor had that the promised extra payment would be made was inconsistent with established doctrine. Nevertheless the Court of Appeal upheld the claim and it is clear that *Roffey* is seen as an instance of the court upholding reasonable expectations.[68]

It seems, then, that in an appropriate case a contractor can reasonably say 'I was unaware of the rule, and my ignorance was reasonable'. On the other hand it is clear that there will be circumstances in which a contractor is entitled to say 'I had a reasonable expectation that the rule book would be applied'. Contractors' expectations are shaped by a range of factors including the terms of the contract itself, but also the negotiations leading to the contract, previous dealings between them and the customs and practices

[68] See Lord Steyn (1997) 113 LQR 433 at p 437–8.

of the particular trade in which they operate.[69] Another factor must, in an appropriate case, be relevant established legal doctrine, which must therefore be a factor in the assessment of what are 'reasonable expectations'.[70] Upholding established doctrine enables contractors to negotiate and contract against a stable backdrop in the expectation that the established rule book will be applied. Where parties have contracted on the basis of established rules, to disapply those rules will frustrate the reasonable expectations of at least one of the parties and deprive them of the benefit of their bargain.[71] This of course is the reason for the strong preference for doctrinal certainty in commercial law.[72]

This is though not to say that expectations are reasonable *merely* because they accord with established doctrine, or, to put it another way, that a party cannot reasonably hold an expectation that is not in accordance with established doctrine.[73] The question then is 'when can a contractor reasonably expect an outcome which is inconsistent with the rule book?' The problem frequently arises in the context of failed negotiations where for one reason or another parties fail to reach final agreement and one then seeks to recover payment for work done or expenses incurred in anticipation of the contract. As *Walford v Miles* shows the general rule of English law, baldly stated, is that until a contract is finalised each party is

[69] Statements made during negotiations may be held to give rise to express contractual undertakings, either as part of the principal contract or by way of subsidiary collateral contract. Previous dealings may lead to the implied incorporation of terms from those dealings. Trade customs and practices may give rise to implied terms. All form part of the relevant background which the court must take into account in interpreting the express terms of the contract. Difficult problems may arise when the parties do not share a common background understanding: see Brownsword, 'After Investors: Interpretation, Expectation and the Implicit Dimension of the New Contextualism' in Campbell, Collins and Wightman (eds), *Implicit Dimensions of Contract* (Oxford, Hart, 2003).

[70] The issue is similar to that raised in cases such as *Mannai Investment Co Ltd v Eagle Star Life Assurance Co Ltd* [1997] 3 All ER 352 and *BCCI v Ali* [2001] 1 All ER 961 where precedent may be considered to be part of the background matrix for the purposes of contract interpretation.

[71] The clash between these two expectations is at the core of the difference between the majority opinions and the dissenting opinion of Lord Wilberforce in *Schuler AG v Wickman Machine Tool Ltd* [1974] AC 325. Applying the language of reasonable expectation to the case the English distributor claimed, and the majority of their Lordships held, in effect, that the distributor reasonably expected that, notwithstanding the use of the word 'condition', the contract would not be terminated on the grounds of a single trivial breach of contract. Lord Wilberforce on the other hand upheld the claim of the German principal that condition had been used as a term of art, and that the principal therefore reasonably expected to be able to exercise the right it had contracted for.

[72] See the comments of Lord Lloyd in *Effort Shipping Co Ltd v Linden Management SA* [1998] 1 All ER 495 at 504.

[73] Note that this is not the same as saying that a party may reasonably expect that the other will not enforce his strict legal rights in accordance with legal doctrine. Past dealings or extra contractual assurances may lead one party to believe, for instance, that the other will not exercise a right to terminate available under the general law. Such assurances may be given effect via the implication of appropriate terms or through doctrines such as estoppel.

free to withdraw from negotiations and to conclude or not conclude a contract as they think fit. But in fact there is a range of devices—including implied contract, tort,[74] restitution[75] and proprietary estoppel[76]—by which English law may protect the reasonable expectations of a person who incurs expenditure during negotiations in this way. That there are, though, limits to this protection, and the relationship between reasonable expectations and legal doctrine, is illustrated by two high profile recent cases.

D. Baird v Marks and Spencer

In *Baird Textile Holdings Ltd v Marks and Spencer plc*,[77] Baird had been a supplier of garments to Marks and Spencer (M&S) for 30 years. The parties had developed a close relationship which several witnesses described as involving a spirit of co-operation close to partnership. M&S encouraged Baird to develop facilities in anticipation of future business. There were regular consultations on strategy, sales, design and other matters. M&S business made up some 30–40 per cent of Baird's annual turnover.[78] However, M&S deliberately refrained from entering into any long term contractual commitment with Baird. Instead orders were placed on a season by season basis. In 1999, without notice, M&S terminated its relationship with Baird. Baird commenced proceedings against M&S claiming that it was precluded by contract and/or estoppel from terminating the relationship. Evidence was given by Sir Richard Greenbury, a former chief executive of M&S, that:

> The special partner relationship which M&S developed with all its suppliers of goods and services was, from its inception some 70 years ago, a cornerstone principle of the company. Furthermore, it was at the very heart of the way we did business with our suppliers and a fundamental part of that philosophy was that M&S was going to carry on doing business with the manufacturers season after season, year after year. Continuity of production into the foreseeable future was the basis of all discussions and negotiations. Indeed it was clearly understood that once a major supplier to M&S always a supplier .... [79]

[74] *Box v Midland Bank Ltd* [1979] 2 Lloyd's Rep 391.
[75] See *William Lacey (Hounslow) Ltd v Davis* [1957] 2 All ER 712; *British Steel Corpn v Cleveland Bridge and Engineering Co Ltd* [1984] 1 All ER 504; *Countrywide Communications Ltd v ICL Pathway Ltd and International Computers Ltd* (1999) (unreported); but cf *Regalian Properties plc v London Dockland Development Corpn* [1995] 1 All ER 1005.
[76] Eg, *Taylor Fashions Ltd v Liverpool Victoria Trustees Co Ltd* [1982] QB 133. See also *Pallant v Morgan* [1953] Ch 43; *Banner Homes Group plc v Luff Developments Ltd* [2000] Ch 372, CA.
[77] [2001] EWCA Civ 274; [2002] 1 All ER (Comm) 737.
[78] The facts of the case bear a striking similarity to the scenario of Bigco and Supplier Inc considered by Allen and Levy, 'The Uses of Ambiguity in Commercial Contracts: On Facilitating Re-Bargaining', above, ch 8.
[79] At para [3].

In short, a major M&S supplier such as Baird could reasonably claim to expect that, in the absence of some complaint about its performance it would continue to be a major supplier of M&S from year to year. This seemed to be recognised even by the principal witness for M&S, its director of procurement, technology and logistics, who described the relationship between M&S and its suppliers variously as involving 'a principle of partnership' and 'symbiotic', within which parties 'were able to trust each other ... and work for mutual benefit'.[80]

Baird might well therefore be said to have a reasonable commercial expectation that it would continue to enjoy M&S's business, and Lord Steyn's comments in *First Energy* about the need to uphold reasonable expectations of honest commercial people were quoted to the court. Nevertheless the claim failed on both grounds. All three members of the Court of Appeal upheld the refusal of the trial judge to imply a contract between the parties. The test for implication of a contract is the same as that for implication of a term into a contract, so that a contract will only be implied if it is necessary to do so.[81] A key factor here was, as Baird acknowledged, that M&S had deliberately refrained from entering into an umbrella contractual relationship in order to preserve commercial flexibility. Implication of a contract would therefore be contrary to M&S's intention. Crucially that intention was known to Baird.[82] Had it been otherwise it might have been possible to argue that objectively M&S's conduct was consistent with the existence of a continuing contract. An even greater obstacle to the implication of a contract, however, was the lack of certainty as to what the terms of that contract would be. As cases such as *Hillas v Arcos* and *Foley v Classique Coaches* demonstrate, any lack of certainty in relation to quantities, prices, delivery dates and so on can be made good by implication of terms based on a standard of reasonableness, what is reasonable being determined by reference to past dealings between the parties. This, however, was not a case like *Hillas v Arcos* where the parties had believed there to be a contract and acted accordingly and the court was prepared to imply appropriate terms to give effect to that contract.[83] The very flexibility of the parties' past relationship made it difficult to adopt that approach. As Mance LJ put it 'the more I have heard and read about the closeness of the parties' commercial co-operation in the past, the less able I have felt to see how its effect could be expressed in terms having any

[80] Para [4].

[81] *The Aramis* [1989] 1 Lloyd's Rep. Treitel argues that the appropriate test in the circumstances of *The Aramis* should have been that not for implication of a term in fact but for implication of a term in law: see 'Bills of Lading and Implied Contracts' [1989] LMCLQ 162. However, the implication of any contract in *Baird* could only have been on the basis of the particular facts of the case.

[82] As appears from para 9.28 of their particulars of claim: see the judgment of Sir Andrew Morritt VC at para 10.

[83] See *Australian Blue Mining Ltd v Hughes* [1963] AC 74.

contractual certainty.'[84] He was particularly concerned that in the event of any future dispute the court would be required to write a 'reasonable' contract for the parties:[85] 'this is not an exercise that the court can or should undertake or indeed which the parties can objectively be taken to have intended'. Significantly, Baird had argued that it would be an implied term of the proposed contract that Baird should deal with M&S in good faith, and *vice versa*. Mance LJ considered the presence of that suggested implied term posed a further obstacle to the implication of the proposed term 'in view of English law's general refusal to recognise any duty of this nature as an implied contractual term.'[86]

The outcome seems harsh and surprising. Emphasis on the individual contracts rather than the overall relationship seems to ignore the economic and commercial reality of the parties' relationship.[87] If there had been a contract between the parties M&S would have no right to terminate other than for serious breach, and even an express right to terminate for breach might be read restrictively.[88] Moreover, had the parties entered into a contract it may well be that obligations requiring the parties to cooperate in good faith could have been implied into the contract.[89] The lack of certainty and completeness in the alleged relationship between M&S and Baird

[84] Para 65.
[85] Para 68.
[86] *Ibid*.
[87] From an economic standpoint the parties might be regarded as having entered into a relational contract. See generally McKendrick, 'The Regulation of Long-term Contracts in English Law' in Beatson and Friedmann (eds) *Good Faith and Fault in Contract Law* (1995). Indeed, it seems that counsel suggested as much: see para 16. That a relationship composed by a series of separate contracts may amount in the law to more than the sum of the individual contracts is recognised by the Commercial Agents (Council Directive) Regulations 1993 SI 1993/3053. Under Regulation 17(6) in the event of termination of the agency the agent is entitled to compensation for the damage he suffers as a result of the termination of his *relations* with the principal. Where there has been a series of contracts between principal and agent reg 17 requires the court to look at the whole of the relationship rather than merely the final contract in the series. See *Duncan Moore v Piretta PTA Ltd* [1999] 1 All ER 174. A further parallel might be with the case law concerned with the question whether casual workers are employed under a global contract of employment: see eg *Airfix Footwear Ltd v Cope* [1978] ICR 1210; *Nethermere St Neots Ltd v Taverner and Gardiner*[1984] ICR 612. Surprisingly these cases were not cited in *Baird*.
[88] See *Schuler v Wickman* (classification of an express term as a 'condition' was held insufficient to permit the principal to terminate where the agent's breach was not serious); *Rice v Great Yarmouth B C* (2000), *The Times* 26 July 2000 (term giving an express right to terminate for breach was held to be subject to an implied restriction that the right would not be exercised in the event of a minor breach).
[89] *Ibid*. There are signs of an increased willingness to recognise implied duties and undertakings in effect requiring good faith from contractors in an appropriate case: see *Phillips Electronique Grand Public BSA v British Sky Broadcasting Ltd* [1995] EMLR 472; *Abu Dhabi National Tanker Co v Product Star Shipping Ltd* [1993] 1 Lloyd's Rep 397: contract term giving a discretion to one of the parties subject to an implied requirement that the discretion must be exercised honestly and in good faith, and not arbitrarily, capriciously or unreasonably; *Paragon Finance Ltd v Nash* [2001] EWCA 1466; [2001] 2 All ER (Comm) 1025: express contractual discretion subject to an implied restriction that it should not be exercised dishonestly, for an improper purpose, capriciously or arbitrarily.

should not have been an insuperable obstacle to the implication of a contract. However, although not cited in the judgments, the influence of *Walford v Miles* can be felt in the court's sceptical approach to the implication of terms requiring good faith and cooperation.

The argument based on estoppel required a development of the law to allow Baird to use estoppel as the basis of a cause of action. The Court recognised that there are instances in which an estoppel may form the basis of a cause of action, most notably in cases of 'proprietary estoppel', but considered that the instant case was one of promissory estoppel or estoppel by convention. The Court held that it was bound by precedent to hold that neither promissory estoppel[90] nor estoppel by convention[91] can form the basis of a cause of action, and that proprietary estoppel is confined to agreements to grant rights over property.[92] The Court was invited to follow the lead of the High Court of Australia in *Waltons Stores (Interstate) Ltd v Maher*,[93] in which during negotiations for the grant of a lease by M to W, M was induced to incur expenditure on the basis of a promise that the proposed lease would be executed, and the court, holding that the promise gave rise to an estoppel which precluded W from denying that a valid lease had been granted, awarded damages to compensate M for its loss incurred in reliance on the assurance. The Court of Appeal rejected that invitation. Not only would it be contrary to established doctrine to allow estoppel to be used as the basis of a cause of action outside the established categories of proprietary estoppel, the obligation created by any estoppel would have to be sufficiently certain to be enforceable. The estoppel claim therefore foundered on the same reef of uncertainty as the implied contract claim.

It is hard not to feel that in *Baird* the claimant's reasonable expectations were not upheld. The Court recognised that Baird had a reasonable commercial expectation that its relationship with M&S would not be summarily terminated. Implied contract and estoppel are two of the principal methods by which English law protects reasonable expectations, and yet neither was able to protect Baird's reasonable commercial expectation. Why is this? Mance LJ expressly recognised that Baird had a *factual* expectation that its relationship with M&S would not be terminated without reasonable notice, but concluded that there was no evidence that Baird had ever 'directed its attention to the legal, as opposed to the factual, position regarding either termination or the parties' 'rights and obligations'.[94] Compare this with the approach in the *Blackpool Aerodrome* case where a contract was implied to align commercial expectation and legal right and

[90] *Combe v Combe* [1951] 2 KB 215, [1951] 1 All ER 767.
[91] *Amalgamated Investment and Property Co Ltd (in liquidation) v Texas Commerce International Bank Ltd* [1982] QB 84, [1981] 3 All ER 577.
[92] *Western Fish Products Ltd v Penwith DC* [1981] 2 All ER 204.
[93] (1988) 164 CLR 387 Aus HC.
[94] See the comments of Mance LJ at para 74.

thus avoid 'an unacceptable discrepancy between the law of contract and the confident assumptions of commercial parties'. The Court cited and applied the *Blackpool* case as authority for the test to be applied when inferring a contract but although it may have followed the letter of the *Blackpool* decision its approach seems a long way from the spirit of that case.

Significantly, with regard to the estoppel claim, Mance LJ suggested that, notwithstanding the doctrinal limitations on the offensive use of estoppel referred to above, on facts similar to those of *Waltons v Maher* the Court of Appeal might be able to reach the same result as the High Court of Australia, treating the estoppel in *Waltons* not as giving rise to a cause of action in itself but as precluding the tenant from raising a 'collateral objection' (the lack of a formally executed lease) to the enforceability of an otherwise complete agreement.[95] It would not have been impossible to construct a similar argument on the facts of the instant case, that M&S were estopped form denying the existence of a global contract. The crucial factor preventing such an argument succeeding here was, it seems, that, *as Baird knew*, M&S had deliberately refrained from entering into a contract in order to retain flexibility in their dealings with Baird. This marks the case out from *Waltons* where the prospective tenant gave a specific assurance that it would execute the lease.[96] To allow Baird's claim to succeed would allow it to assert a legal relationship where M&S had deliberately refused to enter into one. The estoppel case was therefore undeniably weaker than in *Waltons*, but not unarguable. There was no express assurance by M&S that they would not take the point that there was no global contract, but it plainly was arguable that M&S knew, or ought to have known, that Baird understood that there existed a long term relationship between the parties which would be terminable only by reasonable notice, and Mance LJ expressly recognised that Baird understood that there was such a relationship.[97] If M&S understood, or ought reasonably to have understood that that was Baird's understanding—and the evidence from M&S's own witnesses suggested that they did—it would not be impossible to infer a representation to that effect.

E. The Football League Case

It is rare for a contract case to make headlines but *The Football League* case, *Carlton Communications plc and Granada Media plc v The Football League*,[98] made headlines on the front and back pages of national newspapers.

[95] Para 98.
[96] The case is thus closer to the facts of *Austotel Pty Ltd v Franklins Selfserve Pty Ltd* [1989] 16 NSWLR 582. See also *Pridean Ltd v Forest Taverns Ltd* (1996) 75 P & CR 447. In both cases negotiations were being carried on through lawyers, a point emphasised in *Austotel*: see p 585.
[97] Para 74.
[98] [2002] EWHC 1650 (Comm).

The dispute arose out of the collapse of ITV Digital. In 1996 the ITV companies, Carlton and Granada, set up a joint venture subsidiary, ONDigital (later to become ITV Digital, hereinafter referred to as 'Digital') to provide independent television with a terrestrial digital broadcasting platform. The new company began broadcasting in 1998. In 2000 the Nationwide Football League invited tenders for the rights to show live televised coverage of Nationwide League games.[99] Digital bid for those rights and after a period of negotiations lasting over three months a contract between the League and Digital was signed, on 15 June 2000. Under that agreement Digital undertook to pay the League a total of £315 m, including payments of £89.25 m for each of the football seasons 2001/2, 2002/3 and 2003/4, for the exclusive rights to show live coverage of Football League games. The money paid by Digital was to be distributed by the League to the clubs in its three divisions. The contract required Digital and the League to use 'their best endeavours' to execute a long form document within 60 days, the terms of that contract to be 'negotiated' by reference to various earlier documents, including, significantly, Digital's initial bid document of 7 June. No such long form contract was ever concluded, but it was accepted by all parties that the preliminary contract was a legally binding agreement.

This is not the place to comment on the economic wisdom of the Digital contract nor on the morality of football economics. Suffice to say that league clubs entered into extensive financial commitments, mainly for players' salaries, in reliance to a large degree on the income expected to come from the Digital contract. However, Digital was not a commercial success, and its bid to use its football coverage to attract new subscribers was a disaster. In a schedule already saturated with live football coverage the public appetite for football was becoming satiated and live Friday night coverage of Nationwide games was not a particularly attractive proposition (even for supporters of Nationwide teams!). By autumn 2001 Digital was in difficulty and in March 2002 it was placed into administration.

It was clear that Digital would not be able to honour its payment obligations under its contract with the League and that as a result many League clubs would face financial difficulties. The League claimed that Digital's parent companies, Carlton and Granada, had guaranteed Digital's liabilities under its contract with the League. No formal guarantee had been executed and neither parent company was a party to the June contract. The League's case was that in its initial bid document of 7 June, Digital had stated that 'ONDigital and its shareholders will guarantee all funding to the Football League outlined in this document.' That, it argued, constituted an offer of a unilateral contract, which it had accepted by entering into the

[99] Although the most glamorous professional football clubs in England play in the Premiership, the bulk of English professional football clubs—over 75 per cent—play in the Nationwide League.

contract of 15 June with Digital. The parent companies put forward four lines of defence:-

(a) the initial bid was not, properly construed, an offer of a unilateral contract, being made in a proposal which was expressed to be 'subject to contract', and therefore incapable of acceptance; and if it was an offer it had not been accepted;
(b) in any case, the initial bid was made by Digital, which had no authority to act on behalf of the parent companies; even therefore if the bid was an offer of a guarantee it was not binding on Carlton and Granada;
(c) if the initial bid was properly to be construed as a guarantee there was no written memorandum of it as required by the Statute of Frauds, 1677;
(d) if there were a binding guarantee, it would not enable the League to recover damages from the parent companies for an anticipatory breach of contract by Digital.

We are concerned with the first three of these issues which, to some extent, overlap.

The opening words of Langley J's analysis of the first issue were ominous for the League:

> It is an unpromising start for a party who seeks to rely on a guarantee by third parties of obligations involving £315 m entered into by another party that the only reference to a guarantee is to be found in one short sentence of a document produced by the supposed primary obligor in the course of a negotiating process which was 'subject to contract' and the only subsequent effective or binding contract is one agreed by the primary obligor which was on different terms and contains no guarantee nor a reference to one save. In … very oblique terms … It is all the more unpromising when the relevant negotiations are conducted in a major commercial context between two companies with the benefit of the professional advice of experienced managers and lawyers.[100]

Significantly, the director of Digital responsible for preparing the initial bid conceded that the words used could be read as a shareholder guarantee, and Langley J observed that 'in my judgment not only they plainly can but would normally be'.[101] Nevertheless he concluded that the language of the initial bid document could not be construed as an offer of a unilateral contract of guarantee. Having pointed out some peculiarities in the wording of

[100] Para 49.
[101] Para 57.

the document, Langley J observed that it would be unreasonable for the League to rely on it as a guarantee:

> in no normal commercial negotiation do I think a party would be content to rely on such a statement from such a source as providing the security of shareholder guarantees of sums of £240 m.[102]

The crucial factor was the use of the 'subject to contract' rubric, which deprived the statement of any contractual effect. According to authority that rubric remained in force until execution of a formal contract and despite the revised bid which formed the basis of the 15 June contract being expressly stated to be a bid 'on the basis of our initial bid made … on the 7th of June.' The 'subject to contract' rubric was also a key factor in relation to the defence argument that any guarantee was unenforceable for want of a written memorandum complying with the Statute of Frauds. Langley J held that he was bound by the Court of Appeal's decision in *Tiverton Estates Ltd v Wearwell Ltd*[103] to hold that a written memorandum to satisfy the statute must acknowledge the existence of the contract and that a document expressed to be 'subject to contract' will not suffice because it denies the existence of any contract.

Carlton and Granada's third contention was that if an offer had been made by Digital it was made without their authority. There was clear evidence that Digital regularly consulted the parent companies with regard to the progress of its negotiations with the League, and the board of Digital included directors of both parent companies. It is of course trite law that a company has a separate legal personality from its shareholders, and Langley J cited *Salomon v Salomon & Co* and quoted from the judgment of Kerr LJ in *JH Rayner (Mincing Lane) Ltd v Department of Trade and Industry*[104] as authority for the proposition that Digital could not be regarded as the agent of its parent companies merely by virtue of their being its shareholders. There is however relatively little consideration of the question whether, on the facts of the case, the parents had expressly or impliedly authorised Digital to offer guarantees on their behalf and even less of the more pertinent question whether in the circumstances Digital could be said to have apparent authority to offer such guarantees. Indeed, only one short paragraph of the judgement, comprising some seven lines, is given over to the consideration of the possibility of apparent authority, which, it is submitted, should have been the key issue in the case.

[102] Para 59.
[103] [1975] ch 146.
[104] [1989] ch 72 at p 188.

Langley J therefore rejected the League's case on all three grounds.[105] The question for present purposes is whether the result, and the rules which lead to it, can really be said to be in accordance with the League's reasonable commercial expectations? The television companies had opened negotiations with an offer of a parent company guarantee. Had a formal contract been finalised, as anticipated in the contract of 15 June it would in all probability have dealt with the point one way or the other. The absence of a formal guarantee from the short form 15 June contract was however inconclusive. It may well be, of course, that the matter of the guarantee never weighed with the League at all, or that it had no effect on the League's thinking at the time the contract with Digital was signed, and only became a live issue later when Digital's financial difficulties became known.[106] But that is a matter of hypothesis. Let us assume that the League thought that a guarantee of Digital's liabilities was being offered. Was that expectation reasonable? The corporate ownership of Digital was a matter of common public knowledge. Apparent authority arises as a type of estoppel.[107] It is therefore concerned with reasonable expectations, albeit that they must be created by the alleged apparent principal. Was it reasonable for the League to suppose that in making its initial offer Digital was acting with the knowledge and authority of its parent companies, bearing in mind, in particular, the involvement in Digital's management of senior figures from the management of the parents? In *First Energy* it was found that the bank's Manchester manager had apparent authority to make statements on behalf of the bank because that reflected the reasonable commercial expectation of the customer which dealt with the bank through the manager. The bank held the manager out as having authority to make statements by appointing him to his position. Indeed, experience and anecdotal evidence suggests that ordinary people, including ordinary business people, are probably not aware of the effect of corporate personality and the *Salomon* doctrine.

If it was reasonable to believe that any offer of a guarantee was being made on behalf of the parents, the League's expectation that the debt was guaranteed was only unreasonable because of the lack of compliance with the Statute of Frauds and, essentially, because the 'offer' of a guarantee was made 'subject to contract'. *Tiverton Estates v Wearwell*, on which Langley J relied, was concerned with s 40 of the Law of Property Act 1925 which applied to contracts for the sale of land prior to 1989 and required them to be evidenced by a memorandum in writing.[108] Section 40 had its origins in

[105] Although he found that had a guarantee been given it would have extended to the damages liability arising from Digital's anticipatory breach of contract.
[106] Certain passages in the judgment, notably at paras 44–48, suggest that this was indeed the case.
[107] *Rama Corpn v Proved Tin and General Investment Ltd* [1955] 2 QB 147, [1952] 1 All ER 554.
[108] The relevant legislation is now s 2 of the Law of Property (Miscellaneous Provisions) Act 1989.

the Statute of Frauds. However, *Tiverton* is only one of several conflicting Court of Appeal decisions to consider whether the requirements of s 40 could be satisfied by a document expressed to be 'subject to contract', and the Court of Appeal has twice, once before[109] and once after[110] *Tiverton*, expressed the view that a document need not acknowledge the existence of a contract in order to serve as a written memorandum for the purposes of the statute. Thus, although the widely held view is that *Tiverton* is probably correct, the position is by no means clear. Given the existence of four conflicting Court of Appeal authorities on the meaning and effect of 'subject to contract', can we reasonably expect commercial people to understand all the subtle nuances of the law on the subject? In any case, the use of the 'subject to contract' rubric originated in negotiations for the sale of land, which are normally carried out by solicitors or other professionals. The rationale for the Court of Appeal's decision on the meaning of 'subject to contract' in *Tiverton* was that the expression is well known to and understood by conveyancers who regularly use it. In short, to uphold its established technical meaning would accord with the reasonable expectations of those using it. But can we assume that it is equally well known to ordinary commercial people?[111] They might perhaps be expected to appreciate its effect as a warning that no binding contract has yet been concluded, but would they realise that when a formal agreement is concluded, referring back to terms set out in earlier correspondence, that those terms would not have contractual effect because the correspondence was 'subject to contract'? And is it really reasonable to expect ordinary commercial people to be fully aware of the subtleties of a statute passed in 1677, requiring as it does an appreciation of the difference between a guarantee and an indemnity and much other technical learning?[111a]

[109] *Griffiths v Young* [1970] ch 675.

[110] *Daulia v Four Millbank Nominees Ltd* [1978] ch 231.

[111] The course of negotiations between Digital and the League is described in detail by Langley J. It appears that although both parties had legal advisers, they played little, if any, direct role in the negotiations. Negotiations on behalf of the League were largely handled by a specialist media rights consultant, although the League's solicitor did attend the meeting on 15 June at which the short form contract was signed. The use of 'subject to contract' is not confined to property transactions: see *London & Regional Investments Ltd v TBI plc, Belfast International Airport Ltd* [2002] EWCA Civ 355. Other phrases may have the same effect in other contexts: eg, 'subject to details' in relation to charterparties: see *The Junior K* [1988] 2 Lloyd's Rep 583; *The CPC Galatia* [1994] 1 Lloyd's Rep 68; *Manatee Towing Co and Coastal Tug & Barge Inc v Oceanbulk Maritime SA and Laura Maritime Inc (The Bay Ridge)* [1999] 2 Lloyd's Rep 227.

[111a] In *Actionstrength Ltd v International Glass Engineering IN.GL.EN SpA* [2003] UKHL 17 the House of Lords seems to have recognised that the application of the Statute of Frauds to contracts of guarantee may defeat the commercial expectations of contractors. Their Lordships held that a party who had given an oral guarantee was not estopped from relying on the lack of a written memorandum complying with the Statute by the mere fact of having given the oral guarantee but seem to have accepted that an estoppel could have been raised if the guarantor had represented that it would not rely on the lack of a memorandum or that it would confirm the guarantee in writing: see especially the comments of Lord Bingham at paras 8 and 9 and of Lord Walker at paras 51 and 52.

F. Two Comparisons

The decisions in *Baird* and the *Football League* case share some similarities with the cases on letters of comfort. The leading English authority is *Kleinwort Benson Ltd v Malaysia Mining Corpn*[112] in which the defendant company issued a letter of comfort in favour of the plaintiff bank which was contemplating a loan to the defendant's subsidiary. In the letter the defendant indicated that

> it is our policy to ensure that the business of the subsidiary is at all times in a position to meet its liabilities to you under the [loan agreement].

The case was argued on the basis that such letters of comfort are not intended to be legally binding, and that the defendants therefore incurred no obligation to the plaintiffs. The Court of Appeal's finding in favour of the defendant is understandable on the facts of the case, where the defendants had already refused to give a formal guarantee. To hold the comfort letter binding would effectively have given the plaintiffs the guarantee the defendants had refused to give. But the Court decided the case on the basis that the language of the comfort letter expressed only a statement of present fact rather than any promissory warranty: it described the defendants' current policy but gave no promise that policy would be maintained. At first this looks very like legal sophistry and linguistic hair-splitting, the type of literalism from which the law of contract has resiled in the interpretation of contracts.[113] Would the bank have been alert to the precise linguistic niceties of the defendants' letter? There was at that time no reported decision on the effect of a letter of comfort; an expectation that the letter comprised a legally binding undertaking was therefore not contrary to any established rule.[114] However, the case for protection of the bank's expectations in *Kleinwort Benson* is weaker than that for protection of the expectations of Baird or the Football League. In *Kleinwort Benson* the parent company had explicitly refused to give a guarantee before providing the comfort letter, whereas in the *Football League* case the claim was based on the fact that a guarantee had in fact been offered and that offer had never been withdrawn. *Kleinwort Benson* in fact appears closer to *Baird*, where

[112] [1988] 1 All ER 714, [1988] 1 WLR 799 (Hirst J); revsd [1989] 1 All ER 785, [1989] 1 WLR 379, CA.

[113] Lord Steyn offers as an example of pedantic literalism the story of the tyrant Temures who promised the garrison of Sebastia that if they surrendered to him, no blood would be shed, and who, when the garrison surrendered, kept his promise, quite literally, by having the garrison buried alive: (1997) 113 LQR 433 at 441.

[114] The leading practitioner text, Wood: *Law and Practice of International Finance* (1980), cited by Staughton J in *Chemco Leasing SpA v Redifusion plc* (1985), was at the time equivocal about the status of comfort letters, observing that there was no hard and fast rule as to their legal effect.

the crucial factor was M&S' deliberate refusal to enter into a long term contract. However, the essence of Baird's case was that explicit refusal was overridden by the informal assurances, inferred from the totality of M&S behaviour, that the business relationship would continue which generated an expectation which the Court of Appeal recognised was reasonable. The language of the comfort letter in *Kleinwort Benson* was not strong enough to create an expectation in light of the parent's prior refusal of a guarantee. And, significantly, the comfort letter in *Kleinwort Benson* was drafted by the bank itself. It could hardly complain if the words used were ambiguous. Indeed, in light of that fact *Kleinwort Benson* is fully reconcilable with a principle of protecting reasonable expectations. To have allowed the bank to claim that the letter gave rise to a legally binding undertaking in the absence of clear language to that effect would have undermined the legitimate expectation of the parent company.[115]

At the other end of the contract spectrum we may contrast the *Football League* decision with the decision of the Court of Appeal in *Carlyle Finance Ltd v Pallas Finance Ltd*.[116] In *Carlyle* a consumer entered into negotiations to purchase a car by means of a conditional sale agreement. As lawyers well know, such arrangements often involve a tri-partite relationship, with the car dealer selling the car to a finance company which in turn supplies it under a conditional sale agreement to the consumer. Negotiations were, of course, conducted by a car dealer on behalf of the finance company. The finance company approved the proposed transaction and sent the dealer a cheque for the price, together with a conditional sale agreement for completion by the consumer. The proposed conditional sale agreement would be a regulated agreement under the Consumer Credit Act 1974, and it was therefore necessary for the agreement to be executed by both parties in order to comply with the formal requirements of the Act, on failure of which it would be improperly executed, with dire consequences for the creditor.[117] The agreement sent to the dealer was worded as an offer by the consumer to purchase the car, and therefore required completion by the finance company in order to be properly executed. Nevertheless, the finance company authorised the dealer to release the car to the consumer and the dealer allowed the consumer to take possession of the car once the consumer had signed the agreement. The Court of Appeal held that a contract came into being at that moment: the dealer's conduct in allowing the consumer to take possession of the car amounted to acceptance of the consumer's offer (in the conditional sale agreement) to buy it.

[115] A parallel might perhaps be drawn with the 'red hand' cases on the incorporation of unusual terms into contracts, such as *Interfoto Picture Library Ltd v Stiletto Visual Programmes Ltd* [1989] QB 433, [1988] 1 All ER 348, CA; *AEG (UK) Ltd v Logic Resource Ltd* [1995] *CCH Commercial Law Reports* 265.
[116] [1999] 1 Comm 659.
[117] Consumer Credit Act 1974 ss 65, 127.

It is well established of course that an offer is to be interpreted in the way in which it would be understood by a reasonable person in the position of the offeree. The finance company argued that its agent's behaviour in releasing the car could not be interpreted by the consumer as acceptance of his offer because acceptance in that way would not satisfy the requirements of the 1974 Act for creation of a properly executed agreement. The Court of Appeal rejected that contention, however. There was no reason to suppose that a consumer should be aware of the requirements of the 1974 Act.[118]

It must be correct that we cannot reasonably expect a consumer to be aware of a highly technical and complex body of law such as that comprised in the 1974 Act and related regulations, only fully appreciated by specialists. The rationale of *The Football League* case however, is that businessmen are expected to be aware of the law, or, at least, that parties to a large commercial deal can reasonably be expected either to know the law or (more probably) to take legal advice. In the passage from his judgment quoted earlier Langley J observed that the League/Digital negotiations were 'conducted in *a major commercial context* between two companies with the benefit of the professional advice of experienced managers and lawyers.'[119] There is, though, no evidence in the decision that lawyers were involved in the negotiations or preparation of documentation on either side, although it is clear that both sides had the benefit of legal advice.[120]

Where contractors can reasonably be expected to know the law or to have taken legal advice this preference for doctrine is reasonable enough.[121] Indeed, if A negotiating with B reasonably believes that B has taken legal advice, A may reasonably assume that B's expectations are in line with the established legal rule—always assuming that the rule itself is clear. But is it reasonable to expect negotiating parties to take, or even to have access to, legal advice? Such evidence as there is suggests that in fact business prefers not to involve lawyers, especially in the negotiation of contracts. This is especially true of small businesses and may even apply in relatively high

[118] It is perhaps worth noting that the case actually involved a title dispute between two finance houses. The defendant company having bought the car sought to rely on the exception to the nemo dat rule in s 25 of the Sale of Goods Act 1979 (buyer in possession). In order to succeed it therefore had to establish that the consumer was in possession of the car as a person who had bought or agreed to buy it. There were therefore other imperatives in play favouring the finding that there was a binding contract between finance house and consumer.

[119] Para 49.

[120] See also the comments of Hobhouse J in *EE Caledonia Ltd v Orbit Valve Co Europe* [1993] 4 All ER 165 at 173: 'It ... has to be borne in mind that commercial contracts are drafted by parties with access to legal advice and in the context of established legal principles as reflected in the decisions of the courts. Principles of certainty, and indeed justice, require that contracts be construed in accordance with the established principles.' See also his comments at p 177.

[121] A similar issue arises in some of the construction cases. For instance in *BCCI v Ali* [2001] 1 All ER 961 it may have been relevant to Lord Hoffmann's preference for the interpretation of the settlement agreement established by precedent that the employee had received advice from ACAS prior to signing it.

value negotiations. In particular, where no formal written contract is anticipated it will often be the case that lawyers play no role at all in contract formation. Would it be reasonable, for instance, to have expected Baird to have taken advice on the effects of their dealings with Marks and Spencer?

V. CONCLUSION

It is generally accepted that the function of commercial law is to facilitate commercial activity. The law of contract will, on the whole, best do that if it reflects and gives effect to the reasonable expectations of honest commercial people. If there is a mismatch between contract doctrine and commercial expectation one must then yield and be modified to fit the other. All experience suggests that legal doctrine is largely ineffective in shaping business behaviour. If one is to yield, in the absence of overriding policy considerations, it must then be doctrine which yields to practice. There is no value to upholding doctrine for its own sake. This may require us to apply doctrine flexibly, to modify doctrine, for instance by relaxing established rules, to use other doctrines to evade an established rule, or, in some cases, to overturn the established rule.

This supposes, however, that doctrine is in conflict with a reasonable expectation, and that an expectation may be reasonable despite being inconsistent with doctrine. The question then is, can we reasonably expect contractors to know the law? For if we can, we can reasonably expect contractors to act in accordance with it, and an expectation which runs counter to doctrine will, *ex hypothesi*, be unreasonable. Where contracts are made in a setting in which contractors are, or are reasonably expected to be, experienced in the particular market, typically, perhaps, where industry standard forms are widely used, as for instance in the shipping and commodity markets, not only may contractors reasonably be expected to be familiar with the law, but the reasonable expectation of contractors may well be that the established rule will be upheld and applied, and contractors may be taken to contract on that basis. Beyond that we may reasonably expect contractors to be aware of doctrine where they have, or may reasonably be expected to have, legal or (perhaps) other professional advice. Hence the legalistic approach traditionally taken to property contracts is entirely justified. But where contractors cannot reasonably be expected to know the law or to have professional advice, doctrine can realistically play no part in the shaping of expectations. Cases such as *Carlyle* recognise—correctly—that consumers cannot be expected to be aware of detailed technical legal rules. The law must, however, recognise that even in a business setting many contracts are made without the benefit of legal advice. Indeed, lawyers are often seen by business as an obstacle to deal making. As the *Baird* and *Football League* cases demonstrate, even high value transactions are often made without legal advice. To insist on a rigid application of technical doctrine in such cases is in effect to require businesses to take legal advice when contracting, and is unrealistic.

The courts must therefore be flexible in their application of doctrine and be prepared to relax it where doctrine conflicts with reasonable commercial expectation. To a large degree this is already done. Lord Steyn concluded his *Law Quarterly Review* article by observing that:

> in a more formalistic era courts sometimes neglected to consider the reason for a rule. But formalism is receding. Modern judges usually have well in mind the reason for a rule and in a contract case that means approaching the case from the point of view of the reasonable expectations of the parties.[122]

Study of the cases shows that to a large degree English courts are willing to be flexible in the application of doctrine, in a way which may be said to take account of the reasonable expectations of the parties. And yet, as the *Baird* and *Football League* decisions show, sometimes doctrine prevails over what may seem to be reasonable expectation. In both cases it was recognised that the parties had expectations and, seemingly, that they were reasonable and yet those expectations went unprotected, trumped, it would seem, by doctrine on the grounds that the parties (or their advisers) were sufficiently sophisticated reasonably to be expected to know the rules. This is not to say that commercial expectations do not have value in themselves. As counsel observed in a Canadian letter of comfort case, a parent company giving a comfort letter might, when 'push comes to shove' honour its undertaking 'for any or all of the 'non-legal' commercial considerations of reputation, fear of adverse publicity, higher future borrowing costs and a myriad other reasons and possibilities'.[123] But as the *Football League* case shows, if the financial stakes are high enough a company may think that damage to its commercial reputation is likely to cost it less than honouring its non-binding undertaking.

The law of contract performs several functions. One of them is to provide a set of rules for the curial resolution of contractual disputes. However, that is only part of its function. It is recognised that the law should discourage litigation, especially in commercial matters, and the law must therefore provide a framework within which disputes can be resolved without reference to the courts. But contracting parties want, so far as possible, to minimise the risk of disputes arising by providing for potential difficulties in their contracts. In order to do so they must be able to predict the outcomes of cases. A system of precedent allows them so to do, but it will only do so

[122] (1997) 113 LQR 433 at 442.
[123] *Toronto Dominion Bank v Trustee of Leigh Instruments Ltd* (1998) 81 ACWS 3d 117 at para 18.

if the reasons for decisions are clearly articulated. We might choose to express and apply contract doctrine with a greater or lesser degree of rigidity. From the point of view of resolving disputes through litigation it probably matters little whether we adopt a rigid or a flexible approach, although it is to be anticipated that a 'flexible' approach might tend to produce results which are more likely to be considered 'fair'. Rigid doctrine, rigidly applied, will produce greater predictability and will therefore be more efficient in assisting the extra-curial resolution of disputes. For the same reasons it will also be a more effective predictive tool at the planning stage. We might therefore think that a rigid approach to doctrine would be the most desirable from the view point of efficiency. It needs to be borne in mind, however, that the great majority of contracts, including business contracts, are made without the involvement of lawyers. Therefore, unless doctrine is comprehensible to business people it is unlikely to provide much of a guide at the contract planning stage where there is no lawyer involvement. Of the alternative approaches, rigid doctrine, flexibly applied will be wholly inefficient as a guide to the resolution of disputes or the drafting of contracts, for it promises one thing and does another. Flexible doctrine, rigidly applied would appear to be a nonsense. That leaves flexible doctrine, flexibly applied whereby doctrine can be applied more or less rigidly according to the circumstances of the case. Where lawyers are involved or dealings take place in a legalistic context we might expect a stricter approach than where parties contract in a less formal setting without legal advice. The obvious objection to this approach is that it lacks certainty. But the value of certainty is that it facilitates planning. Whilst this approach will be somewhat less predictable for legal professionals than an approach based on rigid application of doctrine, provided we know what factors will guide its application there is no reason why this approach should be unacceptably less predictable. And if we adopt as our guiding principle that the law should reflect and uphold the reasonable expectations of honest commercial people this approach should provide a reasonable degree of predictability at the dispute resolution stage and, because the application of doctrine will accord with *commercial* expectations, also at the planning stage even where lawyers are not involved. Experience shows that a measure of uncertainty can be tolerated; indeed, as Allen and Levy argue elsewhere in this collection, it may even produce more efficient results.[124] In any case certainty may be valued in some markets more than in others. Whilst a high level of predictability, and therefore a rigid adherence to doctrine, may be desirable and appropriate in commodity markets, in other areas of contracting a more flexible approach may be required. The emphasis on doctrinal certainty is emphasised in classical contract doctrine, as reflected in the classic texts, with their emphasis on 'rules' derived from

[124] See ch 8, above.

reported decisions in decided cases, which in turn emphasise a certain type of contract. But 'commercial law' is not only about commodity deals and charterparties. The needs of smaller businesses, for whom contracting may be an altogether less formal business, must also be provided for. Contract law cannot be one size fits all product if it is to meet the needs of all contractors. And as the *Baird* and *Football Legaue* cases show, even a rigid adherence to doctrine cannot deter litigation if the financial stakes are high enough.

If our law is to serve its primary purpose of facilitating commercial activity it is, however, not enough that the law should uphold reasonable expectations. The law must be accessible and comprehensible to the community it serves and, so far as possible, without the intervention of a legal priesthood. Commercial law should serve the commercial community. Technical law, which encourages drafting tricks and technical legal arguments does no service to the commercial community. Too much of our commercial law is lawyers' law, inaccessible in statute, case and textbook. That may be acceptable for large businesses but it needs to be recognised that different businesses have different needs and expectations. Too often references to 'commercial practice' and 'the needs of the commercial community' refer only to the needs of practitioners in legal commercial practice and to the needs of one particular stratum of the commercial community. One example is the emphasis—sometimes over emphasis—on certainty in the law. What business needs from the law is *clarity*. If the reasons for decisions are not clearly articulated then they cannot serve as a guide to future decision making. So if recognition of the broad principles which underlie our doctrinal rules allows us more easily to articulate the real reasons for decisions then we should not be afraid or ashamed to recognise and articulate those principles. To a very large degree there already is flexibility in the application of doctrine. The rules governing contract formation are often less rigid than their description in the books would suggest. If the books are to serve their purpose as a guide to the resolution of disputes and, by backward extrapolation from the results of decided cases, to the preparation of contracts and contracting procedures, they must describe the law accurately. Exaggerated claims of doctrinal rigidity create misplaced expectations. The significance of cases such as *Trentham*, *First Energy* and *Blackpool Aerodrome* is that they not only uphold commercial expectations, but explicitly refer to the need to do so. The increased willingness of our judges to recognise and articulate the reasons for their decisions is a positive development. It behoves those of us who write the books to do the same and recognise the deeper underlying principles which guide the development and application of doctrine.