Exploring Contract Law

Edited by
Jason W Neyers
Richard Bronaugh
Stephen G A Pitel

Exploring Contract Law

Edited by

Jason W Neyers
Richard Bronaugh and
Stephen G A Pitel

OXFORD AND PORTLAND, OREGON
2009

Published in North America (US and Canada) by
Hart Publishing
c/o International Specialized Book Services
920 NE 58th Avenue, Suite 300
Portland, OR 97213–3786
USA
Tel: +1 503 287 3093 or toll-free: (1) 800 944 6190
Fax: +1 503 280 8832
E-mail: orders@isbs.com
Website: http://www.isbs.com

Hart Publishing Ltd, 16C Worcester Place, Oxford, OX1 2JW
Telephone: +44 (0)1865 517530 Fax: +44 (0)1865 510710
E-mail: mail@hartpub.co.uk
Website: http://www.hartpub.co.uk

British Library Cataloguing in Publication Data
Data Available

ISBN: 978-1-84113-906-7

Typeset by Columns Design Ltd, Reading
Printed and bound in Great Britain by
TJ International Ltd, Padstow, Cornwall

EXPLORING CONTRACT LAW

In this book, leading scholars from Australia, Canada, Hong Kong, New Zealand, Singapore, the United Kingdom and the United States deal with important theoretical and practical issues in the law of contract and closely-related areas of private law. The chapters analyse developments in the law of estoppel, mistake, undue influence, the interpretation of contracts, assignment, exclusion clauses and damages. The chapters also address more theoretical issues such as discerning the limits of contract law, the role of principle in the development of contract doctrine and the morality of promising. With its rich scope of contributors and topics, *Exploring Contract Law* will be highly useful to lawyers, judges and academics across the common law world.

Foreword

The chapters in this volume are the product of a symposium titled 'Exploring Contract Law' held at the University of Western Ontario Faculty of Law in January 2008. The sessions featured good intellectual punch-ups between those participants who see law as an adjectival study focused on the work of the courts and commerce, and others who dedicate their careers to more philosophical musings about legal concepts. Both approaches are helpful. Problem solving without a sound philosophical basis risks palm tree justice. Theory without a nod to practice risks irrelevance. Those in the first group would likely agree with Karl Llewellyn's definition of law as what legal people do:

> This doing of something about disputes, this doing of it reasonably, is the business of law. And the people who have the doing in charge, whether they be judges or sheriffs or clerks or jailers or lawyers, are officials of the law. *What these officials do about disputes is, to my mind the law itself*.[1]

However, chapters in this volume by those in the second group show them to be in determined pursuit of what might be called a pure theory of contract law. At the symposium, one of the participants conjured up (in support of the argument that contracts are not necessarily promises) an electrical repairman who includes in his wiring contract a rather unlikely clause:

> The commitments expressed herein are exclusively contractual. We intend hereby to bind ourselves contractually to make the payments and to perform the acts specified, *but we do not intend to bind ourselves morally to do so: these are contractual undertakings, not promises* (emphasis added).

I do not expect to see such contractual provisions anytime soon and the theoreticians probably have no expectation that we will, or even much interest in whether we do. Such writings constitute 'an exercise in logic, not in life,'[2] and derive their nourishment from the writings of other academics rather than judges, even the sort of peripatetic judges who turn up at symposia like this to listen, learn and inwardly digest.

What is intensely enjoyable about these sorts of confrontations is the enthusiasm with which the participants attack one another. For example, John Swan observed at the symposium:

[1] K Llewellyn, *The Bramble Bush: Some Lectures on Law and its Study* (1930) 3 (emphasis in original), recently reprinted as K Llewellyn, *The Bramble Bush: The Classic Lectures on the Law and Law School* (New York, Oxford University Press, 2008).

[2] HJ Laski, *A Grammar of Politics*, 4th edn (London, Allen & Unwin Ltd, 1963) vi.

> From my point of view, I can't see what could possibly be gained by the acceptance of Stephen [Smith's] argument that, for example, the law of damages for breach of contract is not part of the law of contract but should instead be dealt with in a discussion of both contract and tort damages, with tort damages seeming to run from cases of personal injury through negligent misrepresentation to trespass, defamation and beyond.

The 'what's your point' rejoinder was much favoured by most of the participants in both camps most of the time. While the chapters published here are generally more restrained in tone than the symposium itself, together they represent significant and closely-argued contributions to our collective understanding of contract law. We are fortunate to have the work of both the theoreticians and the Llewellynites gathered together here in permanent form.

Justice Ian Binnie
Supreme Court of Canada
28 May 2008

Contents

Contributors

RICK BIGWOOD, LLB, PHD, Professor of Law, University of Auckland

RICHARD BRONAUGH, BA, MS, PHD, BLITT, Professor of Law and Professor Emeritus of Philosophy, University of Western Ontario

MINDY CHEN-WISHART, BA, LLB, LLM, MA, Reader in Contract Law, Merton College, University of Oxford

HELGE DEDEK, First and Second State Examination in Law, LLM, PHD, Assistant Professor of Law, McGill University

GERALD H L FRIDMAN QC, MA, BCL, LLM, FRSC, Professor Emeritus of Law, University of Western Ontario

MARK P GERGEN, BA, JD, Fondren Foundation Centennial Chair for Faculty Excellence, University of Texas

ANDREW S GOLD, BA, JD, Associate Professor of Law, DePaul University College of Law

KELVIN F K LOW, LLB, BCL, Associate Professor of Law, The University of Hong Kong

JASON W NEYERS, BA, LLB, MST, Associate Professor of Law and Cassels Brock LLP Faculty Fellow in Contract Law, University of Western Ontario

STEPHEN G A PITEL, BA, LLB, LLM, PHD, Associate Professor of Law, University of Western Ontario

ANDREW ROBERTSON, LLB, LLM, PHD, Professor of Law, University of Melbourne

STEPHEN A SMITH, BA, LLB, PHD, William Dawson Scholar and Professor of Law, McGill University

ROBERT STEVENS, BA, BCL, Professor of Commercial Law, University College London

ANDREW TETTENBORN, MA, LLB, Bracton Professor of Law, University of Exeter

CHEE HO THAM, LLB, BCL, Associate Professor of Law, Singapore Management University

CATHERINE VALCKE, LLB, LLB, LLM, JSD, Associate Professor of Law, University of Toronto

STEPHEN WADDAMS, BA, MA, PHD, LLB, LLM, SJD, FRSC, University Professor and Goodman/Schipper Professor of Law, University of Toronto

CHARLIE WEBB, BA, LLM, PHD, Lecturer, London School of Economics and Political Science

Introduction

JASON W NEYERS, RICHARD BRONAUGH
AND STEPHEN G A PITEL*

The chapters in this book arose from the 'Exploring Contract Law' symposium held at the University of Western Ontario, Faculty of Law, on 10–11 January 2008. The symposium brought together leading contract scholars, private law theorists and judges from around the common law world. The purpose of the symposium was to have these thinkers examine contract law with fresh eyes—to explore it anew. Thus, the presenters were asked to explore contract law (or its related doctrines) in one of three ways. First, they could (re)explore doctrines that are considered tangential or antiquated. This task was undertaken, for example, in the presentations by Mindy Chen-Wishart and Gerald Fridman. Second, they could explore what appeared to be settled principles in light of recent case law developments. This was done most clearly in the presentations by Robert Stevens, Kelvin Low and Rick Bigwood. Third, they could explore black letter contract law from a theoretical or comparative perspective. Many of the presentations did this in one form or another.

As noted by Justice Binnie in his foreword, collectively these explorers can also be divided into two groups, though not without overlap. Some explorers, particularly the authors of the chapters which open this collection, took a theoretical perspective. Others took a more practical view, concerned with doctrine and its impact on practitioners and the courts. Depending on their own leanings, readers will find that some of the resulting chapters blaze new trails through terrain that feels more familiar to them, while others bushwhack into less charted territory.

Beyond the formal presentations, which have become chapters in this collection, the symposium was enriched by commentary and questions from scholars and theorists such as Peter Benson,[1] Ralph Cunnington,[2] Daniel Markovits,[3] John McCamus,[4] Michael Pratt[5] and John Swan[6] and

* Each of the Faculty of Law, University of Western Ontario.

[1] P Benson (ed), *The Theory of Contract Law: New Essays* (Cambridge, Cambridge University Press, 2001).

[2] R Cunnington and D Saidov (eds), *Contract Damages: Domestic and International Perspectives* (Oxford, Hart Publishing, 2008).

[3] D Markovits, 'Contract and Collaboration' (2004) 113 *Yale Law Journal* 1417.

[4] JD McCamus, *The Law of Contracts* (Toronto, Irwin Law, 2005).

[5] M Pratt, 'Promises, Contracts, and Voluntary Obligations' (2007) 26 *Law and Philosophy* 531.

[6] J Swan, *Canadian Contract Law* (Markham, LexisNexis Butterworths, 2006).

from justices of the Supreme Court of Canada, the British Columbia Court of Appeal and the Court of Appeal for Ontario. For two days, those assembled were treated to a rich diet of scholarship and informal exchange and this collection is the result of that endeavour.

* * *

Stephen Smith's chapter opens the book by seizing on its theme of exploration.[7] He does this not in the way of those seeking out new lands in distant places but in the way of those who 'advance tentative observations' on what is already in place. Thus he wishes to improve the existing map of contract law, that is, to be a logical, local cartographer and not a scout. To this taxonomical end, he argues that many of the rules that are conventionally regarded as contractual, and that, as a consequence, are regularly invoked to support a particular theory of contract law or explained on the basis of such a theory, are not in fact contractual at all. Instead, they should be regarded as belonging to the (hitherto unrecognised) general part of the law of obligations. Smith concludes that failing to pay attention to contract law's special border—which he labels 'vertical'—with the general part can lead 'scholars, judges and lawyers to apply the wrong principles to understand legal rules, to draw the wrong inferences from those rules, and to fail to make appropriate generalisations'.[8]

Helge Dedek's chapter is inspired by the works of Peter Birks in general and Stephen Smith in particular and offers a comparative law perspective on the taxonomic debate in common law scholarship.[9] Of course, the attempt to impose a logical order on the law is the hallmark of civilian, and particularly German, 'legal science' of the nineteenth century. Without seeking to discredit the undertaking, Dedek acknowledges that modern approaches have surpassed nineteenth century formalism by explicitly taking the normative (or, as Smith puts it, 'moral') implications of legal classification into account. However, drawing from experiences with the German civil law, Dedek focuses on the downside of dividing the law of obligations into a 'general' and a 'special' part, namely technical problems regarding the formulation and application of abstract 'general' rules and the problem of the 'hardening of categories,' a phenomenon which could lead to an intellectual rigidity that curtails the argumentative potential of legal discourse. The latter problem is illustrated by comparing how different legal traditions conceptualise the 'reliance interest' in contract damages.

[7] 'The Limits of Contract' ch 1.
[8] *Ibid*, at Part VIII.
[9] 'Border Control: Some Comparative Remarks on the Cartography of Obligations' ch 2.

In his chapter, Stephen Waddams examines the concept of principle—that is, what we mean by calling something a principle—in relation to the doctrine of consideration.[10] Whereas Smith and Dedek are concerned with the general and special borders of contract law, Waddams is more concerned with content and substance than with form and taxonomy. Yet this concern has formal or taxonomical effects relevant to the prior two chapters. Waddams notes that Blackstone did not think of contract as an independently existing field of law but rather thought of it as part of several different areas of the law. Hence it would have no special borders and would be found in several different places on a map. Waddams argues that a historical study cannot establish either (a) the correct meaning of the word 'principle' or (b) the 'correct' rules of contract law. Being unable to find historically a single or fixed content for the doctrine of consideration, Waddams concludes simply that the regular appeal by the courts to the concept of principle has historical significance and shows that neither the usage nor the law (and its taxonomy) is immutable or eternal.

Catherine Valcke's chapter compares two national legal systems and so further complicates the issue of the taxonomy and content of contract.[11] Her aim is twofold: (a) to demonstrate that there are peculiarly different French and English thought structures animating the intellectual life and law of French and English societies, and (b) that one of the purposes of comparative law, as a self-standing academic discipline, is to uncover and compare these different thought structures. The chapter draws parallels between French contract law (subjective intention as it informs issues of mistake and interpretation) and Rousseau's conception of the state, and draws parallels between English contract law (objective intention as it informs the same issues) and Hobbes' conception of the state. Valcke argues that although these conceptions may differ in their social value, they are internally accurate in that each correctly reflects the distinctive self-understandings of contract law in the two nations. Such a conclusion has clear implications for exploring or 'mapping' contract law through form and content.

Andrew Gold examines yet another border issue for contract law, namely the line between a moral agent's understanding of the obligation of promising in ordinary life and promissory obligation as enforced in contract law.[12] Gold addresses the claim, recently made by Seana Shiffrin, that the doctrine of consideration presents a troubling divergence from

[10] 'Principle in Contract Law: the Doctrine of Consideration' ch 3.

[11] 'Contractual Interpretation at Common Law and Civil Law: An Exercise in Comparative Legal Rhetoric' ch 4.

[12] 'Consideration and the Morality of Promising' ch 5.

promissory morality.[13] Noting that Shiffrin considers matters only from the promisor's point of view, Gold takes up the position of the moral promisee. Here Gold finds that promissory morality and contract law do not diverge. He affirms that when a gratuitous promise is broken, the moral promisee gains a right personally to rebuke the promisor, but no more—neither in morality nor in law. But when the promise received by the promisee is part of an agreement grounded in an exchange then a stronger right to performance is acquired by the promisee, both morally and legally. In fact, this emergent moral right is to a remedy from the promisor in case of breach and justifies creating (as substitute for self-help) the legal power for coercion through the state—something well beyond rebuke. The existence of contractual consideration thus serves, Gold argues, to justify an augmented moral right, a right that cannot exist in this way when the promise has been gratuitous. This conclusion is contrary to Shiffrin's divergence thesis about the consideration doctrine. The quid pro quo requirement of the bilateral contract should not present an embarrassment for the agent seeking to live a moral life.

Gold's chapter makes the moral and legal right to a remedy, typically damages, a consequence of a promisee's having provided consideration in a bilateral contract to the promise breaker. Yet Smith, in the first chapter in this volume, would not regard damages as part of contract law proper. To understand damages, he holds, one must look up to what is more general in law because 'the duty to pay damages is not uniquely a response to breach of contract'.[14] Still, one cannot deny (and Smith does not deny) that damage rules must be understood in order to appreciate the practice of enforcing contracts.

One turns then to the chapter by Charlie Webb, who asks what justifies an award of damages for breach of contract when performance clearly is what a party contracts for.[15] Webb observes that: 'To obtain performance is one thing; to receive a sum of money to make up for the losses caused by not obtaining performance is something different.'[16] For example, if the right to performance is taken seriously, then it would seem that specific performance should be the natural remedy. Nonetheless, the occasion for the performance may be past or its reclamation may demand undue supervision by the court. What then? Webb claims that this contractual right 'can, sometimes, be effectuated through an award … which the claimant uses to purchase an equivalent "performance" from an alternative

[13] S Shiffrin, 'The Divergence of Contract and Promise' (2007) 120 *Harvard Law Review* 708.

[14] 'The Limits of Contract' ch 1 at text following n 21.

[15] 'Justifying Damages' ch 6.

[16] *Ibid*, at text following n 10.

source'.[17] The aim is to find remedies which as much as possible adhere to the 'special' domain of contract law. Any damages award that goes beyond simulating performance, Webb claims, requires reflection on 'the norms and ideals which shape and justify the law'.[18] Thus, Webb would seem to be responsive to Smith's distinction between 'special' and 'general' since it appears that once one leaves the realm of specific performance and damages which simulate performance, it appears as if one must depart from the 'specific' and pass through into the 'general' area of the law of obligations.

In the second damages chapter,[19] Robert Stevens assesses whether the recent House of Lords decision in *Golden Strait Corporation v Nippon Yusen Kubishika Kaisha (The Golden Victory)* was rightly decided.[20] His ultimate conclusion is that the result is defensible if one differentiates between damages representing the right to performance (Webb's focus) and those awarded for consequential losses. Using examples drawn from a wide range of contracts (such as sale, carriage and construction) Stevens argues that this differentiation is necessary since: (i) the rules for the assessment of these two types of damages (such as mitigation, remoteness, timing of assessment and ancillary benefits) are quite different, and (ii) it is the only coherent way to justify otherwise inexplicable differences in damages quantification. Thus Stevens argues that although damages for breach of contract are commonly thought to be awarded only to compensate the claimant for loss or, more rarely, to strip the defendant of a gain, 'general' damages are commonly quantified by reference to the value of the contract right which the defendant has infringed, despite the absence of any consequential loss. Moreover, he claims that once awards of 'substitutive' damages are accepted, it is doubtful whether there is significant authority supporting the award of a remedy assessed by reference to the defendant's gain.

Having considered the general border of contract law through the law of remedies, there is the special border still to be examined. How does contract interact with or separate itself from other areas of law on the same plane of a legal structure? This is the place where every field of law, while 'special,' can be interactive with others. The next five chapters explore the special or horizontal border.

In Anglo-Australian law, it is unclear whether proprietary and equitable estoppel form part of the law of contract, part of the law of wrongs, or a separate part of the law of obligations. Adopting Birks' taxonomy that organises private law according to various rights-creating events, Andrew

[17] *Ibid*, at Part VII.
[18] *Ibid*.
[19] 'Damages and the Right to Performance: A *Golden Victory* or Not?' ch 7.
[20] [2007] 2 AC 353 (HL).

Robertson argues that these estoppels cannot be seen as part of the law of wrongs because it is possible to identify a series of events which give rise to primary rights that are recognised by the law prior to and independent of any infringement of those rights.[21] Robertson then argues that these doctrines cannot be seen as part of the law of contract either because a promise is neither necessary nor sufficient to establish liability. As a result, Robertson concludes that proprietary and equitable estoppel must be classified as *sui generis* and therefore belong in Birks' category of 'other' events. They are specially (horizontally) separate from contract law and indeed not interactive.

Gerald Fridman's chapter is about excessive crossing of the border between tort and contract.[22] By investigating the tort of inducing breach of contract, he assesses the protection that the reasonable expectations of the contractual parties receive in the law of tort. Fridman's argument is that the foundational, yet extremely controversial, decision in *Lumley v Gye*,[23] was not supported by precedent and is a significant example of judicial activism. He criticises the courts, old and new, for being too willing to use tort law to protect contracts. In the second half of the chapter, Fridman argues, in opposition to the recent judgment of the High Court of Australia in *Zhu v Treasurer of New South Wales*,[24] that the defence of justification on moral grounds should be significantly expanded to better protect those who induce breaches of contracts.

In his contribution, Mark Gergen examines American law on the effectiveness of agreements that absolve an actor from liability for misleading another.[25] This is to consider the interaction of contract and the law of wrongs. Although American courts are divided on the issue, Gergen contends that the courts that refuse to enforce provisions which exculpate from fraud are correct. He argues that innocent parties can be adequately protected from baseless accusations by rules requiring that fraud be pled with specificity and proven by clear and convincing evidence. In relation to inadvertent misrepresentation, Gergen argues that terms that bar a contract claim on a representation should also bar claims in rescission, restitution, or negligent misstatement since this is a contract law issue rather than one of equity, restitution or tort. Turning to negligent misstatement, Gergen observes that an exculpatory agreement will determine the existence and scope of an actor's duty of care since invited reliance is a *sine qua non* of liability. In this important respect, Gergen argues that negligent

[21] 'Estoppels and Rights-Creating Events: Beyond Wrongs and Promises' ch 8.
[22] '*Lumley v Gye* and the (Over?)Protection of Contracts' ch 9.
[23] (1853) 2 El & Bl 216, 118 ER 749 (QB).
[24] (2004) 218 CLR 530 (HCA).
[25] 'Contracting Out of Liability for Deceit, Inadvertent Misrepresentation and Negligent Misstatement' ch 10.

misstatement resembles contract, in particular the doctrines of promissory estoppel and third beneficiary in American law, and is unlike the general tort of negligence.

The rights under a contract can be, and frequently are, assigned to a third party. Less frequently, a contracting party can create a trust of the rights under a contract in favour of a third party beneficiary. In his chapter, Andrew Tettenborn compares and contrasts these two approaches to transfer.[26] The distinction is particularly important in light of two relatively recent English cases, *Don King Productions Inc v Warren*[27] and *Barbados Trust Co v Bank of Zambia*.[28] These cases take the position that the two approaches are different, and if one is unavailable—for example because of an express prohibition on assignment—the other remains open. Tettenborn challenges these decisions as failing to reflect commercial reality and creating a distinction lacking any real difference. Tettenborn's approach takes a broad and strong view of clauses in a contract which prohibit the benefits under the contract from being assigned to a third party. On that view, such clauses should prohibit the creation of a trust of the contract's benefits.

In contrast, Chee Ho Tham's chapter argues for a minimalist approach to such 'third party' clauses.[29] He supports the analysis in *Don King* and is critical of *Linden Gardens Trust Ltd v Lenesta Sludge Disposals Ltd*, a case in which the House of Lords relied on an anti-assignment clause to invalidate a purported equitable assignment.[30] In Tham's view, equitable assignment should be understood as involving only a transfer of the right to grant a release and its corollary, the right to decide to bring legal proceedings on the chose in action. Anti-assignment clauses thus cannot prevent an assignor from effectively transferring, in equity, his or her right to grant such a release. Tham's claim is that these clauses are less effective than the courts currently think they are.

The final group of chapters deals with the concrete doctrines of common law that seek to protect parties in situations of information or power imbalance. This time the border is not between special fields of law horizontally considered or as fields captured by general, vertically appreciated, principles of law. The issue here is about the relationship between persons involved in making an agreement within the bargaining confines of contract formation. In his contribution,[31] Kelvin Low examines the English Court of Appeal's recent judgment in *Great Peace Shipping Ltd v Tsavliris*

[26] 'Assignments, Trusts, Property and Obligations' ch 11.
[27] [2000] Ch 291 (CA) [*Don King*].
[28] [2007] EWCA Civ 148, [2007] 1 Lloyd's Rep 495 (CA).
[29] 'The Nature of Equitable Assignment and Anti-Assignment Clauses' ch 12.
[30] [1994] 1 AC 85 (HL).
[31] 'Coming to Terms with *The Great Peace* in Common Mistake' ch 13.

Salvage,[32] which reformulated the law of common mistake by overruling *Solle v Butcher*.[33] Low finds it surprising that *The Great Peace* has found itself under such sharp criticism—a development which demonstrates a mellowing of views of both academics and judges over Denning LJ's excesses in *Solle v Butcher*. In fact, instead of criticising *The Great Peace*, he claims that it has wrought two principal improvements to the law: the abolition of a dual jurisdictional approach to cases of common mistake which caused an unnecessary amplification of uncertainty, and the clear severance of the doctrine of common mistake from the implied condition precedent theory. Low contends that these improvements allow the common law scope for developing its remedies of mistake beyond the blunt instrument of declaring the contract void *ab initio*.

Mindy Chen-Wishart's chapter examines the foundational and controversial case of *Smith v Hughes*[34] and the objective theory of contract formation.[35] She concludes that situations that are said to represent examples of subjectivity trumping the objective approach are straightforward applications of objectivity when a context-specific view is taken. In fact, she argues that when this view of objectivity is applied, there is no need, indeed no room, for a resort to subjectivity. Her chapter also attempts to stabilise the language of 'mistake', 'defective consent' and 'void' and make distinctions between contract formation and vitiation. Her argument is that this distinction explains why known non-correspondence of any term prevents contract formation, while mistaken assumptions must be shared and fundamental to void a contract. This distinction also allows us to map the related areas of rectification, *non est factum*, mistaken identity and misrepresentation.

In the final chapter of the book, Rick Bigwood examines the nature of undue influence, a doctrine frequently employed by vulnerable parties to set aside contracts.[36] He explains the different categories of cases, in particular highlighting the differences between cases where the undue influence is established on the facts of the case and cases where the undue influence is rooted in the relationship between the parties. He analyses in detail the fiduciary elements on which this second category of cases has been based, and is critical of the House of Lords for having lost sight of these important underpinnings in its more recent decisions, notably *National Westminster Bank plc v Morgan*[37] and *Royal Bank of Scotland*

[32] [2003] QB 679 (CA) [*The Great Peace*].
[33] [1950] 1 KB 671 (CA).
[34] (1871) LR 6 QB 597.
[35] 'Contractual Mistake and Intention in Formation and Vitiation: the Oxymoron of *Smith v Hughes*' ch 14.
[36] 'From *Morgan* to *Etridge*: Tracing the (Dis)Integration of Undue Influence in the United Kingdom' ch 15.
[37] [1985] AC 686 (HL).

plc v Etridge (No 2).[38] In Bigwood's view, the court's move away from strict fiduciary regulation in relational undue influence cases is a retrograde step.

* * *

Both the symposium and this book have benefited from generous support and assistance. We are indebted to the faculty, staff and students at Western Law, particularly Dean Holloway, Acting Dean Brown, Associate Deans Edgar and Huscroft, and our conference coordinator, Michelle Bothwell, for their enthusiastic support of the symposium. We are also grateful for the financial support provided by Western's Research Promotion Fund, which financed the operating costs of the symposium, and by the Foundation for Legal Research, which allowed us to have the editorial assistance of two talented law students, Jean-Michel Corbeil and Carrie Ann Miller. We are also indebted to the law firm of Cohen Highley LLP for graciously sponsoring the symposium banquet and to Justice Ian Binnie for providing the foreword to the collection.

We hope that you will enjoy reading the chapters in this book as much as we enjoyed hearing them presented as papers at the symposium. We are confident that they will be helpful guides to anyone who is interested in exploring contract law.

[38] [2002] 2 AC 773 (HL).

1

The Limits of Contract

STEPHEN A SMITH

THE THEME OF this book brings to mind two images. The first is that of a legal scholar who, like the explorers of old, goes in search of new or little-understood contract law terrain. The other image is that of a scholar who advances tentative observations or arguments about contract law. In form and substance, this article is an exploration in the second sense. Focusing on well-known legal rules, it offers tentative observations and arguments about the law of contract and related areas of private law. But it is the first image that provides the article's inspiration. This article seeks to provide guideposts for those who explore (knowingly or not) the outer reaches of the law of contract.

I. EXPLORERS AND BORDERS

Contract law explorers are sometimes content to do no more than describe what they have found on their journeys. But most such scholars also pursue one or the other (or both) of two further aims. The first is to apply ideas and concepts drawn from the broader law of contract to the new or little-understood rules in the hope that these ideas and concepts will assist in understanding those rules. The second aim is to apply what they have learned about the new or little-understood rules to the broader law in the hope that these lessons will shed light on the broader area. In practice, most scholars pursue both aims, moving back and forth between particular contract law rules and broader contract law ideas and concepts.

Neither of these two activities can be undertaken without a preliminary idea of the scope or terrain of the law of contract. Legal exploration, by its nature, occurs near borders. If exploratory scholars are not careful they may leave the territory of contract without realising it. Terminological confusion is a typical consequence. But the more serious consequence of scholars failing to notice that they have left the terrain of contract is that they may attempt to understand cross-border law using inappropriate

concepts (that is, contract law concepts) or may use that cross-border law to draw inappropriate conclusions about contract law—or may do both these things. The main reason for paying attention to contract law's borders is to ensure that the tasks of understanding particular rules and drawing conclusions from those rules are done properly. To be sure, contract law's borders are not fixed. They are the product of ongoing reflection about the nature of contracts, which is informed by exploratory scholarship. One reason scholars explore is to test, inform and, where necessary, revise their views about the borders between contract and other legal fields. But that complex task cannot be undertaken without an idea, however preliminary, of what those borders might be and on what basis they have been constructed.

II. HORIZONTAL AND VERTICAL BORDERS

The law of obligations is conventionally divided into sub-categories organised around particular kinds of obligations. Thus, the sub-category of contract law is associated with the obligation to perform a contract, while the sub-category of unjust enrichment is associated with the obligation to reverse an unjust enrichment. The third main sub-category, the law of torts, though not identified with a single obligation, is generally understood to be comprised of a number of sub-categories, each of which is associated with a particular obligation. Thus, torts textbooks typically discuss, inter alia, the law of trespass (the obligation not to commit a trespass), the law of defamation (the obligation not to defame others), the law of nuisance (the obligation not to interfere with others' use of their property), and the law of damages (the obligation to pay sums to those whom one has defamed or to those against whom one has committed a trespass, committed a nuisance, etc).

It is not clear if these obligation-specific categories are meant to exhaust obligations law (so far as I am aware the issue has never been addressed directly), but the absence of other recognised categories suggests that this is the working assumption of most common law scholars. Yet if this is the assumption, it raises an immediate question: How, if at all, is the idea that there is such a thing as a 'law of obligations' reflected in this scheme? If the law of obligations is a meaningful category, then the various obligations that comprise it must have something in common. Further, if the law of obligations is a meaningful *legal* category, these common elements should be reflected in the law. The very concept of a law of obligations appears to assume the existence of rules that cut across or otherwise tie together obligation-specific categories such as contract, trespass, and so on. It appears to assume, in other words, that the law of obligations contains not just obligation-specific rules but also general rules.

The concept of a 'general part' of the law of obligations, though familiar to civilian lawyers,[1] has not traditionally been part of the common law lawyer's vocabulary. It seems clear, however, that there exist rules within the common law that cut across or otherwise tie together the obligation-specific categories. These rules appear to be of three main kinds. The first group is comprised of rules that set common pre-conditions for different obligations. An example is the set of rules governing the creation and existence of 'artificial legal persons' such as registered corporations or charities. These rules are applied by courts in both contractual and non-contractual settings. Only legal persons can be bound by contracts. Equally, only legal persons—as defined by the same set of rules—can be held responsible for injuring others. Secondly, there are rules that set common 'enforcement' conditions for different obligations. Limitation periods, for example, are often common in this sense. Thus in the United Kingdom both actions founded on torts and actions founded on contracts expire six years after the cause of action accrued. Finally, there are entire categories of obligations that are dependent upon or derived from other obligations. The obligation to pay damages, for example, arises if and only if another obligation—which may be contractual or non-contractual—has been breached.

Understood in this way, the law of obligations may usefully be organised along the lines of a schema that until now has been applied (in the common law anyway) primarily within the field of criminal law. Criminal law scholars conventionally divide the rules that constitute the criminal law into those belonging to the 'special' part of criminal law and those belonging to the 'general' part.[2] The former is comprised of rules that identify particular offences (for example, theft, murder and assault), while the latter includes rules that are applied generally, that is to say, to more than one kind of specific offence (for example, the defences of duress and insanity and the concept of *mens rea*). The equivalent distinction within obligations law is between rules used to identify particular primary obligations (for example, the obligation to perform a contract, the obligation not to trespass and the obligation to tell the truth) and those that are applied to obligations generally or at least to more than one kind of obligation. 'Primary' obligations, as understood here, are obligations that are not dependent on or otherwise derived from other obligations. The

[1] This distinction is especially well known in German law and scholarship: see, eg, R Zimmermann, *The Law of Obligations* (Oxford, Clarendon Press, 1996) 29–31; K Zweigert and H Kotz, *An Introduction to Comparative Law*, 3rd edn, trans T Weir (Oxford, Clarendon Press, 1998) 146–7.

[2] See, eg, S Shute and AP Simister (eds), *Criminal Law Theory: Doctrines of the General Part* (New York, Oxford University Press, 2002); RA Duff and SP Green (eds), *Defining Crimes: Essays on the Special Part of the Criminal Law* (New York, Oxford University Press, 2005).

qualification is necessary to ensure that the law of damages—which is clearly concerned with a specific type of obligation—is distinguished from those obligations on which it is dependent (for example, the obligation to perform a contract, the obligation not to trespass, etc).[3] Though different in important ways from the legal personality rules and rules on limitation periods (neither of which serve to identify a distinctive obligation), the damages rules are appropriately regarded as general because, like the personality and limitation rules, they cut across or otherwise tie together various specific obligations. This view of damages is defended in more detail below. For the moment, it is sufficient to accept that obligations to pay damages are dependent upon, and therefore structurally different from, obligations to perform contracts, to not trespass, to not commit a nuisance and so on.

The distinction between the special and the general parts of obligations law suggests that it may be helpful to think about the organisation of obligations law using the civilian image of a pyramid of rules. The rules on the top (which belong to the general part of obligations law) apply to more than one kind of primary obligation, while the rules on the bottom (which belong to the various special parts of obligations law) apply to individual instances of primary obligations. It also follows from this way of understanding obligations law that contract law's scope should be demarcated in terms of two borders. The first, which I will call the 'horizontal' border, separates contract law rules from the other sets of primary obligation-specific rules that occupy the bottom of the obligations pyramid. It thus separates contract law rules from rules identifying other obligations that, while arising from different events[4] and having different contents, operate on the same plane as contractual obligations. As just noted, that plane, in broad terms, is occupied by other primary obligations such as the obligation not to trespass, the obligation not to damage another's property, the obligation not to defame and so on. It does not include obligations that aim to cure, repair or otherwise respond to a rights-infringement, such as the obligation to pay damages.[5]

[3] Obligations to make restitution are intentionally not mentioned in this paragraph nor discussed elsewhere in this article because their classification raises special, and as yet unresolved, issues. In a forthcoming essay, I argue, contrary to the conventional view, that restitutionary duties are structurally closer to duties to pay damages than to duties to perform contracts, to not commit trespasses, etc: S Smith, 'Torts and Unjust Enrichments, Damages and Restitution' in R Chambers, P Mitchell and J Penner (eds), *The Philosophical Foundations of Unjust Enrichment Law* (Oxford, Oxford University Press, forthcoming 2009). If this is correct, restitutionary duties belong within the general part of obligations law.

[4] Here as elsewhere in this article the word 'event' includes status events such as coming of age or entering the jurisdiction (which then give rise to obligations to respect others' persons, property, etc). The more accurate alternative of describing obligations as arising 'from events and from status' is grammatically awkward.

[5] On the classification of restitutionary duties, see above n 3.

The second border, which I will call contract's 'vertical' border, sets the boundary between contract law and the general part of the law of obligations. This border is vertical because it sets the boundary between rules at the top and bottom of the pyramid; specifically it distinguishes contract law rules (and other primary-obligation rules) from rules that are located above them in the sense that they qualify these rules or regulate the consequences of breaching the obligations to which these rules give rise.

This article focuses on contract law's vertical border. It addresses an imbalance. Legal scholars have devoted considerable attention to contract law's horizontal border. There is an extensive literature on the nature of a contract (for example, whether contracts are promises, reliance-inducing statements, transfers or something else), on the difference between contract and other fields (for example, misrepresentation law and trusts), and on the classification of things like estoppel, the trust, bailment, gratuitous undertakings, promises under seal and simultaneous exchanges—all of these are concerned with the location of contract's horizontal border.[6] In contrast, contract law's vertical border has received little attention. As already noted, the very idea that there exists a general part to the law of obligations is almost unknown in the common law.[7]

This neglect has had significant consequences. Admittedly, the conclusion that a particular rule, say a rule on duress, is within the general, rather than the special, part of obligations law does not tell us anything about the content or purpose of the rule other than that both should be in some sense general, that is, applicable to more than one kind of primary obligation. By contrast, the conclusion that a particular rule or set of rules are within this or that aspect of the special part of obligations law often has immediate practical implications. It may determine, for example, the standard of care expected of a defendant or the scope of the defendant's responsibility for breaching the obligation. But the neglect of contract law's vertical border has had serious consequences for our understanding of the law and, in the long run, for the law's development. It has led to misunderstandings about the nature of contract law and about the nature of certain rules wrongly assumed to be contract law rules. It has also impeded our understanding of other legal rules. The common law has failed to understand or even to recognise certain legal categories because their members have been wrongly assumed to lie within contract law (or tort law, etc). Rules with similar foundations and aims have been applied and studied in isolation from one another.

[6] I discuss some of this literature in S Smith, *Contract Theory* (New York, Oxford University Press, 2004) chs 3, 5 and 7, and S Smith, *Atiyah's Introduction to the Law of Contract,* 6th edn (New York, Oxford University Press, 2005) 28–35, 65–93.

[7] A notable recent exception is P Cane, 'The General/Special Distinction in Criminal Law, Tort Law and Legal Theory' (2007) 26 *Law and Philosophy* 465.

These are strong claims, made here without examples. I make no pretense of developing or defending them in full. The article's primary aims are more modest: to introduce the idea of contract law's vertical border, to suggest an approach for locating this border and, finally, to identify a few categories of rules that, while commonly described as contract law rules, prima facie appear to lie outside contract's vertical border.

III. CLASSIFICATORY CRITERIA: TWO QUESTIONS

Locating contract's vertical border is an exercise in legal taxonomy. Like other taxonomic exercises in law, it is undertaken with the aim of making the rules under examination more intelligible by showing how they are like or unlike other legal rules. This task may be undertaken in different ways; the appropriate criteria depend on the classifier's specific interest. For lawyers, judges, legal scholars and others with a general, rather than a specialised, interest in the law, the criteria that are normally applied are essentially moral criteria. For example, the familiar distinctions between tort, contract, and unjust enrichment are based on the conviction that these labels correspond, at least in broad outline, with morally significant distinctions (for example, between self-imposed and externally imposed obligations, and, within the latter, between obligations not to harm and obligations to return benefits). Legal scholars who question the existence or scope of these categories generally do so not on the basis that they reject moral criteria, but on the basis that the legal categories, at least as currently understood, fail to reflect significant moral categories. This approach is appropriate for anyone with a general interest in the law, as the law is defined in large part by the fact that it purports to create moral obligations.[8] The most fundamental general question about the law is whether it succeeds in this aim, that is to say, whether the law is justified. Trying to classify legal rules on the basis of moral distinctions helps in answering this question.

The relevant moral distinctions will differ, however, depending on the kinds of rules that we are trying to distinguish between. Thus when attempting to distinguish contract law from its neighbours, it makes a difference which border—the horizontal or the vertical—is under scrutiny. The point of fixing the horizontal border, recall, is to distinguish between contract-creating events and other kinds of primary obligation-creating events. There are different views as to how this distinction should be drawn, but nearly all rely on variants (or combinations) of three broad views about the nature of a contract, each of which locates the source of

[8] See, eg, J Raz, *Ethics in the Public Domain* (Oxford, Clarendon Press, 1994) 210–20.

contractual obligations in a different morally distinctive event: (1) contracts are promises (or agreements); (2) contracts are transfers (of already-existing rights); and (3) contracts are reliance-inducing statements.[9] In some cases, it is possible to conclude that a particular rule lies inside or outside contract's horizontal border without deciding between these views. Thus, a sufficient explanation for why the rules governing trespass to land are non-contractual is that on any plausible account of the nature of a contract-creating event (promise, transfer, reliance-inducing statement or something similar), the trespass rules are unrelated to that event. In other cases, however, it is necessary to be more precise about the nature of a contract-creating event in order to classify the rule. Thus, to determine whether the rules governing executed gifts are contractual it is necessary first to determine whether contracts are promises, reliance-inducing statements, transfers or something else. Only if contracts are transfers is it plausible to conclude that executed gifts qualify as contractual events. Similarly, to determine if some or all of the rules dealing with misrepresentation or with estoppel might be classified as contractual, it is necessary to determine, inter alia, if the mere inducement of reliance (without a promise) can create or modify a contractual obligation.

In the case of contract's vertical border, where the issue is whether the rule in question belongs to the special or general part of the law, the aim is to distinguish between rules that identify the existence, meaning or consequences that flow from a *particular* obligation-creating event and rules that identify the existence, meaning or consequences that flow from obligation-creating events *generally* (or at least in more than one case). The relevant moral distinction is therefore between rules that should apply specifically to contractual obligations and those that should apply more generally. The importance of asking whether it makes sense, morally, for the rule to be applied generally cannot be overstated.[10] Different obligations are often governed by rules that, while similar in broad outline, differ in their details. These differences may be nothing more than accidental by-products of a process in which the law is developed through adjudication of individual disputes. Alternatively, they may reflect differences in the application, but not the substance, of a common principle. Or, finally, they may reflect a difference in underlying foundations. Only in the last case do

[9] See Smith (2004), above n 6, at ch 3.

[10] The importance of these questions is recognised in the literature on the parallel distinction between the general and the special parts of the criminal law, though most scholars have tended to emphasise one or the other question, not both. Thus, Glanville Williams in *Criminal Law: The General Part* (London, Stevens and Sons Ltd, 1953) holds that the distinctive feature of rules in the general part is that they are applied to multiple offences, while Michael Moore in *Placing Blame: A General Theory of the Criminal Law* (Oxford, Clarendon Press, 1997) argues that their distinctive feature is found in their distinctive moral purpose.

the differences preclude classifying the rule as general. But we cannot tell which one is the correct interpretation without an idea of what kinds of rules it makes sense to suppose are general and what kinds it makes sense to suppose are specific.

Equally, however, it would be a mistake to conclude that a rule can be classified as within the general part solely on the basis that this makes moral sense. The law that we are trying to render more intelligible is the law that we actually have. Any classification of rules must take into account how those rules are understood and distinguished by the people who actually use them. The consideration doctrine, for example, is difficult to explain on the basis of any of the leading views about the nature of a contract. It is not a part of the concept of a promise, transfer, reliance-inducing statement, or any other plausible candidate for the basic contract-creating event. It is indisputable, however, that the consideration doctrine is part of the law of contract and not a part of the general law. The reason is found in the positive law: the consideration doctrine, whatever its purpose, is consistently applied to—and only to—contractual obligations. Another way of putting this point is that it is almost certain that the basic contract-creating event is not a 'simple' or 'unitary' event such as a promise, transfer and so on. Rather, the event is something more like 'a promise given in exchange for consideration'. We know this, inter alia, because the consideration rule is applied to contractual obligations *and* because it is not applied to other obligations.

To summarise, the answers to two questions must be taken into account—no stronger term is warranted—when trying to determine if a particular rule is within the law of contract or within the general part of the law of obligations: (1) is the rule applied to resolve not just contractual but also non-contractual disputes? (2) is the rule of a kind that it makes sense, morally, to suppose should be applied to obligations generally as opposed merely to contractual obligations?

IV. AN OVERVIEW

This article's substantive argument is that a significant number of rules discussed in articles and textbooks ostensibly devoted to contract law are part of the general law rather than part of contract law. Before discussing these rules, however, it may be useful to say a few words about what *is* contractual. Briefly, and allowing that the terms are subject to varying interpretations, the law of contract comprises what would be labeled in most Anglo-American textbooks as the rules dealing with offer and acceptance, consideration and intention to create legal relations; as well as those concerned with the incorporation, implication, and interpretation of contract terms (the latter group to include at least some of the rules on

mistake, frustration, termination and unenforceable contracts). These rules are contractual because, on the one hand, as a matter of positive law they are applied only to contractual disputes, and, on the other hand, because it makes sense to regard them as contractual. It makes sense because they define the existence or meaning of the basic contract-creating event (for example, a promise, transfer etc) *or* they qualify the existence or meaning of this (and only this) event. The rules on offer and acceptance, for example, identify the basic contract-creating event, while the rules on consideration and intention to create legal relations qualify the test for the existence of that and only that event. Similarly, the rules on incorporation and interpretation of terms define the meaning of the event (for example, the meaning of a promise), while the rules on implication of terms (specifically implied-in-law terms) qualify that meaning and only that meaning.

Topics that are discussed in nearly every contract textbook such as damages, specific performance, duress and undue influence are not included in this list. Further, in so far as the excluded topics are indeed non-contractual, the reason, in many cases, must be that they lie outside contract's vertical border. The rules on damages, specific performance, duress and undue influence, for example, are regularly applied by courts when they are resolving contractual disputes. If these rules are not contractual, the explanation cannot be that they are a part of misrepresentation law or negligence law, that is to say, that they belong to another special part of obligations law. Rather, the explanation must be that they are within the general part of obligations law, that is to say, that they apply both to contractual and non-contractual obligations. It is not possible within this article to discuss all or even most of the rules that fit this description. Instead, I will focus on three categories of rules that receive significant attention in nearly all contract textbooks and that are regularly invoked by contract scholars either to support a particular view of contract law or to be illuminated by that view, but which belong, I will argue, to the general part of obligations law.

To avoid misunderstandings, my argument is not that the rules examined below should never be discussed in contract textbooks, courses, articles and so on. The lawyers, judges and students who are the main readers of contract law literature typically approach contract law questions from a problem-solving perspective. They want to solve real-life (or hypothetical real-life) contract problems, and to do that they need to be familiar with all the rules that are applied to contractual disputes, whatever their origin. Further, we can often learn important things about how courts understand contracts by seeing how general rules are applied to contractual disputes. I will say more about this below. For the moment, it is sufficient to observe that while there are good reasons to discuss general rules in contract textbooks and so on, this should be done self-consciously. There is always

the danger, as mentioned earlier, of drawing the wrong implications from these rules or using the wrong tools to examine them. In the case of textbooks, at least, it is not clear to me if the issues discussed in this article are even raised. Every author must of course make decisions about what to include and what to exclude. Some of the rules that are normally excluded from contract textbooks are part of the general law; for example, the rules on legal personality are not usually discussed in contract textbooks. But it is not clear what, if any, criteria are employed in such decisions. The dozen or so contract textbooks on my bookshelves all contain explicit references to the borders between contract and tort, contract and unjust enrichment and so on, but there is no mention of anything resembling the distinction between contract law and the general law which is the focus of this article.

V. RIGHTS ARISING FROM NON-PERFORMANCE (1): THE LAW OF (ORDINARY) DAMAGES

Every common law contracts textbook discusses the rules governing damages for breach of contract, usually in considerable detail. Often regarded as the most famous article in English on contract law, Fuller and Purdue's 'The Reliance Interest in Contract Damages'[11] focuses, as the title states, on damages rules. Contract courses in common law jurisdictions frequently begin with a discussion of damages. When I studied contract law, the first cases I read—*Hadley v Baxendale*[12] and *Peevyhouse v Garland Coal & Mining Co*[13]—were both about damages. But the rules governing contract damages are not contract law rules. Nor—though the point is a fine one—do they belong elsewhere within the special part of obligations law. Although the damages rules govern a specific type of obligation—to pay damages—this obligation operates on a different plane than the primary obligations to performs contracts, not to trespass, not to injure another and so on. The obligation to pay damages arises on the breach of any primary obligation. It is therefore a part of the general law.

Damages orders come in different shapes and sizes.[14] But in the typical case an order to pay damages compels the defendant to do either or both of two things. The first is to pay a sum equivalent to the cost of repairing, replacing or purchasing a substitute for whatever property or service was

[11] (1936) 46 *Yale Law Journal* 52.
[12] (1854) 9 Exch 341, 156 ER 145.
[13] 382 P 2d 109 (Okl 1962).
[14] SM Waddams, *The Law of Damages*, 4th edn (Toronto, Canada Law Book, 2004); AM Tettenborn, *The Law of Damages* (London, LexisNexis UK, 2003); S Smith, 'The Law of Damages: Rules for Courts or Rules for Citizens' in R Cunnington and D Saidov (eds), *Contract Damages: Domestic and International Perspectives* (Oxford, Hart Publishing, 2008).

damaged, destroyed, lost, taken or not provided as a direct result of the defendant's breach of duty.[15] The second is to pay a sum equal to any losses that the claimant suffered indirectly as a result of the above breach, for example losses arising from the delay in receiving actual or substitute performance. Whether these are separate duties[16] or different aspects of the same duty[17] is an interesting question. For present purposes the important point is that both duties are distinct from the duty to perform a contract. The latter is a duty to do what you said you would do. Although contracting parties can stipulate that a particular sum must be paid in case of breach, the duty to pay ordinary damages is not grounded in any such agreement.[18] Defendants are required to pay damages not because they agreed to pay damages, but because they breached the contract. The duty to pay arises at the moment of injury or termination and the content of the duty is set by various rules, including those on remoteness, mitigation and so on.[19] When applying these rules it is necessary to determine the content of the defendant's contractual obligation; however, that obligation does not itself determine the amount of damages. The defendant's contractual obligation is to deliver goods, construct a building, provide a service and so on. If this obligation is not met, the defendant is normally liable for the cost of obtaining a substitute. That cost is not set by the contract. Moreover, the defendant is also liable for whatever consequential losses the claimant suffers as a result of late delivery. Again, this sum is not set by the parties' contractual agreement. This is true even where the defendant is fully aware, when entering the contract, of the law on damages. Being aware that one's actions may have certain consequences does not mean that one agrees to those consequences. Contracting parties agree only to the consequences to which they have actually agreed. Thieves do not agree to be punished merely because they were aware, at the time they committed their crimes, of the sanctions for theft.

The conclusion that damages rules are not contractual does not rest, however, solely on the fact that the duty to pay damages does not have the same source as ordinary contractual duties. As mentioned earlier, it is both logically possible and in practice almost certainly the case, that contract law is mixed in the sense that some contractual duties (or limits on those duties) derive from a promise (or reliance-inducing statement, transfer etc),

[15] Where the breach consists merely of late performance, no damages are awarded under this first heading.

[16] As I suggest in S Smith, 'Substitutionary Damages' in R Grantham and C Rickett (eds), *Justifying Remedies in Private Law* (Cambridge, Cambridge University Press, 2008).

[17] As argued by E Weinrib in 'Two Models of Damages' in Grantham and Rickett, *ibid*.

[18] Nor is payment of the stipulated sum the same as payment of damages. A court order to pay a stipulated sum is simply an order to pay a debt; it is not an order to pay damages.

[19] The qualifying rules differ depending on which of the two kinds of damages just described is at issue: see Smith, above n 16.

while others derive from a different source (for example, a general concern for fairness in contracting). A duty to pay damages might be thought to fall within the latter category. Specifically, it might be argued that a duty to pay damages is an implied-in-law contractual term, that is to say, a term that the law implies into a contract unless the parties stipulate otherwise.[20] Courts and legislatures commonly imply non-mandatory terms into contracts. Whether non-mandatory implied-in-law terms are properly regarded as creating contractual as opposed to tort-like duties is a matter for debate,[21] but however they are classified they are not part of the general law. Implied-in-law contractual terms impose specific duties within specific relationships. It might be argued, therefore, that the duty to pay damages is effectively an implied-in-law stipulated damages clause.

The difficulty with this response—and at the same time the reason that damages rules are properly considered part of the general, rather than the special, law—is that the duty to pay damages is not uniquely a response to breach of contract. The duty to pay damages is a general duty in the sense that it arises not just on the breach of contractual duties, but also on the breach of other primary duties, such as duties not to trespass, cause a nuisance, negligently harm others and so on. For both contractual and non-contractual wrongs, the basic measure(s) of damages are—and should be—the same. In each case, the defendant must pay for the cost of 'substitute performance' in the form of repairing or purchasing a substitute for whatever property or service was damaged, destroyed, lost, taken or not provided directly as a result of the defendant's breach and, in addition, must pay a sum equal to any consequential losses that the claimant suffered indirectly as a result of the breach.[22] This is appropriate: if the aim of the damages is to provide substitute performance and compensate losses arising from a failure to perform an obligation, damages should in principle be equally available irrespective of the nature of obligation.[23]

In practice, the actual damages a claimant obtains may differ depending on whether the underlying duty is contractual or non contractual. But with rare exceptions this is not because different principles are applied to

[20] Some economist-lawyers explain damages in precisely this way; see eg, I Ayres and R Gertner 'Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules' (1989) 99 *Yale Law Journal* 87; JS Johnston, 'Strategic Bargaining and the Economic Theory of Contract Default Rules' (1990) 100 *Yale Law Journal* 615.

[21] See Smith (2004), above n 6, at 307–14.

[22] The not uncommon view that the distinction between damages assessed at the level of cost of cure and damages assessed at the value of performance, reveals something important about the nature of contractual obligations, is difficult to maintain once it is recognised that claims for tort damages for the loss or destruction of personal property raise similar issues that are answered in similar ways: see in particular Waddams, above n 14, at 3–82.

[23] Subject to the qualification that the obligation in question is not an obligation to repair or compensate, as in the case of an obligation to pay damages or make restitution: see Smith, above n 3.

contractual and non-contractual breaches, but rather because the application of the general principles underlying the law of damages is sensitive to the nature of the duty that was breached. The contents of duties to pay for substitute performance and duties to compensate for consequential losses are determined, at least in part, by the nature of the duty that is breached. To determine which losses were caused by a breach of duty you need to know what that duty is. If the duty is to produce a certain outcome (as is true of most contracts) then the direct and indirect consequences of failing to produce that outcome will be different from those associated with, say, breaching a duty to take care. It is this difference in application of the same principle, rather than a difference in principles, that explains why the basic measure of damages obtained for breach of contract is often greater than that received, on similar facts, for tort claims.

The same is true of the limitations on damages imposed by the rules governing remoteness and mitigation. Although the basic principles of remoteness and mitigation of loss are the same for contract and tort claims, their application may differ because, again, those principles often require courts to look at the nature and scope of the primary duty. While is easier to say what 'remoteness' is not than what it is (for example, it is not merely about probability), in broad terms remoteness is about 'responsibility for'. Assessments of responsibility for losses are based, in part, on the prior relationship between the victim and the injurer. The relationship between contracting parties is typically—though not always—closer than the relationship between the parties to a tort dispute. The explanation of mitigation is similar. Although the term is used more commonly in contract cases than tort cases (in tort the concept of mitigation is usually subsumed within the broader notion of remoteness), the underlying principle applies equally to tort claims. In both contract and tort, claimants cannot recover for losses that they could have reasonably avoided.[24] What counts as reasonable will often (though not always) differ depending on the nature of the breach. The victim of a contractual breach will often be expected to continue dealing with the wrongdoer.[25] But this merely shows that 'reasonable' must be understood in context. The importance of the remoteness and mitigation rules lies in what they tell us about the law of damages, not what they tell us about the nature of contractual obligations.[26]

[24] This is not to deny that courts sometimes say that there is a difference between how loss is quantified in tort and contract claims (see, eg, *C Czarnikow Ltd (The Heron II) v Koufos* [1969] 1 AC 350 (HL)), though it is unclear how often, if ever, such statements have actually affected the final decision in a case. The common law tradition of treating damages claims as falling into one or another distinct body of law ('the law of damages for breach of contract,' 'the law of damages for tort') has encouraged this kind of thinking.

[25] *Payzu Ltd v Saunders* [1919] 2 KB 581.

[26] Authors who have attempted to derive propositions about the nature of contractual obligations from the rules on remoteness or mitigation include G Gilmore, *The Death of*

This interpretation of damages rules does not deny that damages decisions often provide useful evidence for contract scholars. If damage awards normally reflect the primary duty in the ways explained above, then the amount of damages awarded in any particular case may provide a clue as to court's view of the duty that was breached. But care must be taken when drawing conclusions about contractual obligations from cases on contract damages. We get from damages to contract through a theory of damages, not contract. Only if we know why damages are awarded can we say what a particular award implies about the nature of the primary duty. This point must be stressed because the relationship between damages awards and the duties for whose breach they are awarded is complex. We have already noted that an award of 'ordinary' or non-adjectival damages provide redress in at least two quite different ways. Further complexities arise in cases involving nominal damages, punitive damages, damages for mental distress and other forms of what might be styled 'vindicatory' damages.[27] In all these cases damage awards may well tell us something about the underlying primary duty, but in none of them does the award reflect that duty in a straightforward fashion.[28] The lesson is clear: to say anything meaningful about contract law on the basis of damages awards, it is necessary first to have a theory of why damages are awarded. And to produce such a theory, it is necessary to view damages in their entirety, that is, to view damages for all breaches of primary duties together. The law of contract damages is one part of the general law of damages.

VI. RIGHTS ARISING FROM NON-PERFORMANCE (2): THE LAW OF COURT-ORDERED RIGHTS

Contractual obligations are private obligations in the sense that they are owed to other citizens or to entities that are treated in law as ordinary citizens. Governments make contracts, but in so far as these contracts are governed by ordinary contract law rules (as they normally are) the government is treated as a natural person. It follows that the rules comprising the law of contract are directed fundamentally at citizens: contract law rules tell citizens how they should behave when they interact with other citizens, at least within a certain sphere of activity. But a not

Contract (Columbus, Ohio State University Press, 1974) 49–53; R Danzig, '*Hadley v Baxendale*: A Study in the Industrialization of the Law' (1975) 4 *Journal of Legal Studies* 249; W Bishop, 'The Contract–Tort Boundary and the Economics of Insurance' (1983) 12 *Journal of Legal Studies* 241.

[27] R Stevens, *Torts and Rights* (Oxford, Oxford University Press, 2007) 59–92; D Pearce and R Halson, 'Damages for Breach of Contract: Compensation, Restitution, and Vindication' (2008) 28 *OJLS* 73.

[28] Stevens, *ibid*, at 59–92; Tettenborn, above n 14, at 25–51; Smith, above n 14.

insignificant number of rules found in standard contract textbooks are directed at courts, not citizens. In particular, the duty-imposing rules dealing with contract remedies, save those concerned with quantifying an ordinary damages award (see above), are directed at courts. These rules tell courts how they should behave when making court orders. It is arguable that these rules do not belong to obligations law (or even private law), let alone the law of contract. But if they are a part of obligations law, they are within the general part.

A. Orders to Perform an Existing Duty

Court orders fall into two broad categories.[29] In the first category, the order (merely) confirms an existing duty, as for example when a court orders specific performance of a contract or when it orders a defendant to pay damages for consequential losses. Such orders confirm the defendant's already-existing duty (in the examples just given to perform the contract or to compensate for consequential losses). The *content* of orders in this first category is therefore determined by the rules that govern the content of the relevant pre-existing duty. In some cases, these are contract law rules (as where specific performance is ordered) and in some cases they are not (as where damages are ordered), but they are all private law rules, directed at citizens. Thus, the content of an order to perform a contract is determined by the ordinary contract law rules on the interpretation, incorporation, and implication of contract terms, while the content of an order to pay compensation for consequential losses is governed by the damages law rules dealing with remoteness of loss, mitigation and so on.

The *availability* of specific performance and other orders in this first category is not, however, determined by the rules governing contracts, damages, or any other private law duty. The conclusion that John has a contractual duty to deliver goods to Ann tells us what John should do (deliver goods to Ann), but it tells us nothing about what the courts or any other organ of the state should do if John fails to perform that duty. These are different questions. When a claimant asks a court to make an order against the defendant, the claimant is asking the state to involve itself in what has been, until then, a private matter. There may be good reasons for courts to accede to such requests; this explains why many court orders are available 'as of right'. But such rights are rights against courts, not against the defendant or any other individual. The rules that govern such rights are not a part of the law of contract. Indeed, it is not clear that they are a part

[29] R Zakrzewski, *Remedies Reclassified* (New York, Oxford University Press, 2005) ch 5; Smith, above n 14.

of the law of obligations or even of private law.[30] Although rights to court orders arise from infringements of private rights and although the court orders to which they give rise are themselves a source of private obligations, rights to court order are strictly against the state and, as such, arguably a part of public law.

For present purposes, the final classification of rights to court orders can await another day. What is important is to see that whether they are public, private, or mixed public/private, these are general rights in the sense that they are applied to disputes arising from different kinds of obligations. A court's willingness to order a defendant to perform an existing duty depends in every case on the content of that duty, but it does not depend on whether the duty is contractual or not.[31] Existing duties to pay money, for example, are invariably enforced directly if the duties are due and the claimant so requests, provided that the relevant limitation period has not expired. This is true whether the underlying duty is to pay damages, to pay a contractual debt or to return money paid by mistake. In all these cases, an order to pay the money is available as of right. By contrast, most non-monetary orders are subject to a complex set of pre-conditions, particularly if the relevant duty is a duty to do something as opposed to a duty *not* to do something. Once again, this is true whether the original duty is contractual or not. Thus courts are generally willing to enforce negative duties not to trespass, not to create a nuisance or not to breach a restrictive contractual covenant. Conversely, positive duties, whether in contract or not, are typically enforced only if they are relatively simple and if substitute performance is unavailable.

The generality of the rules governing specific relief is particularly important for contract scholars in light of the considerable attention such rules have traditionally received in (ostensibly) contract law literature. Probably the most famous sentence ever written in English about contractual obligations—Holmes's 'The duty to keep a contract … means … that you must pay damages if you do not keep it,—and nothing else'[32]—is about the rules governing specific performance. The same is true of what is probably the best-known idea in contract scholarship of the last half-century, the theory of 'efficient breach'.[33] Contemporary contracts scholarship continues to devote considerable attention to remedial rules and, in particular, to the rules governing specific performance.[34] As was true of

[30] See Smith, above n 14.

[31] See Smith, above n 16.

[32] 'The Path of the Law' (1897) 10 *Harvard Law Review* 457, 462.

[33] RL Birmingham, 'Breach of Contract, Damage Measures, and Economic Efficiency' (1970) 24 *Rutgers Law Review* 273; JH Barton, 'The Economic Basis of Damages for Breach of Contract' (1972) 1 *Journal of Legal Studies* 277.

[34] See, eg, D Kimel, *From Promise to Contract: Towards a Liberal Theory of Contract* (Portland, Hart Publishing, 2003); SV Shiffrin, 'The Divergence of Contract and Promise'

damages rules, it may be that we can learn something indirectly about contractual obligations by observing how specific performance rules are applied. But even more than in the case of damages rules, the rules on specific performance have a complex relationship to the concept of a contractual obligation. The rules on specific performance stipulate various situations in which a state actor (a court) has the authority to do or not do something. As mentioned already, in so far as claimants have a right to specific performance, it is a right against the state. It is to be expected, therefore, that the rules governing the availability of specific performance (and of other court orders) will reflect concerns that are particular to the state and to the order that the state is contemplating making. The question the court must consider is not whether the defendant had a duty to do what the claimant alleges (that question has already been answered), but whether the court should respond to the failure to perform that duty by invoking its power to order specific performance or, instead, by doing something different. It would not be surprising to find, therefore, that institutional concerns (broadly defined) explain much of the law in this area. In practice, this is exactly what we do find. Without going into detail, the limits on specific performance (and other requests to enforce existing duties) appear to reflect, inter alia, an institutional concern for simple, easily enforced orders, and, in addition, for avoiding orders that might be regarded as involving something akin to servitude. The rules on specific performance therefore tell us a great deal about how courts understand their duties, but very little about how courts understand citizens' duties.

B. Orders to Perform a New Duty

The other category of court orders is comprised of orders that create, by the order itself, new duties. An example is an order to pay punitive damages. Prior to coming to court a defendant is not under a duty to pay punitive damages to the claimant. You cannot punish yourself. The duty to pay punitive damages is created by the order. It is difficult to say with certainty how many other court orders fall into this second category. The most likely candidates appear to be nominal damages, ordinary damages in cases where performance by the defendant was possible and desired when the parties came to court (that is, where the contractual obligation remained in force until the order), and perhaps damages for mental distress and loss of satisfaction.[35] For present purposes, the important point is that

(2007) 120 *Harvard Law Review* 708; A Bagchi, 'Contract v Promise' <http://ssrn.com/abstract=1012150> (2007) (last accessed: 1 November 2008); D Markovits, 'Making and Keeping Contracts' (2006) 92 *Virginia Law Review* 1325.

[35] See Smith, above n 14.

whatever the size of this category, the orders that fall within it provide very limited and indirect evidence of the nature of the underlying obligation. As was true of 'duty-confirming' orders, the availability of 'duty-creating' orders is not determined by contract law principles. It is determined by the general principles that govern when (if ever) it is appropriate for courts to create new duties through court orders. But the distance from the underlying duties is even greater than in the case of duty-confirming orders. Duty-creating orders neither confirm the underlying primary duty (as happens with specific performance orders) nor, with one exception, reflect that duty (as is true of ordinary damages). A court's decision to award punitive or nominal damages, for example, tells us that the court believes a duty was breached, and, in the former case anyway, may tell us something about the court's view of the seriousness of the breach and the defendant's motives, but the content of the award is only loosely connected to the content of the underlying duty. Of the orders mentioned above, only orders to pay damages that are made in lieu of ordering specific performance appear to reflect the content of the underlying duty. In such cases, damages are calculated on the same basis as ordinary damages.

VII. INCAPACITY, DURESS, UNDUE-INFLUENCE AND MISREPRESENTATION: THE LAW OF RESPONSIBILITY

Courts are generally uninterested in the attributes of contracting parties, the circumstances under which they make their contracts or their reasons for contracting. The same rules of offer and acceptance, interpretation, and so on are applied to the old and the young, the rich and the poor and the informed and the uninformed. But there are exceptions. In certain circumstances, an arrangement that would otherwise create a binding contract will not have that effect because of the contracting parties' attributes, circumstances or reasons for entering the contract. Some of the rules that provide for such a result are plausibly explained as either implied-in-fact or implied-in-law contractual terms, and, as such, as belonging to the special part of obligations law, if not contract law itself. This seems the most plausible interpretation of (most of) the rules regarding mistake,[36] frustration and termination for breach.[37] But there are many 'defences' that cannot be explained in this way, in particular those provided for by the rules on incapacity, duress, undue influence and misrepresentation.

The law dealing with incapacity, duress, undue influence and misrepresentation is large and complex. In broad terms, however, there appear to be

[36] Excepting mistake as to 'terms', which is part of the law of offer and acceptance: see Smith (2004), above n 6, at 366–7.

[37] See *ibid*, at 365–74.

two main ideas underlying these rules.[38] The first is that individuals should not be bound to contracts if their ability to consent in a meaningful way was significantly impaired, either generally or with respect to the specific contract in question. Most cases in which incapacity or undue influence is pleaded successfully seem best explained on this basis. The other idea that appears to underlie much of the law in this area is that wrongdoers should not profit from their own wrongdoing (where 'profit' includes obtaining contractual rights). The rules that allow a contract to be set aside for a negligent or intentional misrepresentation seem best explained on this ground. A large number of rules, in particular the rules on duress, appear to be explicable on either of the above grounds. The grounds are, however, distinct. A contract may be set aside for incapacity regardless of the fault or even knowledge of the party seeking enforcement. This is a purely 'consent-based' explanation. Conversely, a contract entered on the basis of a misrepresentation or an unlawful pressure emanating from a third party (as opposed to the other contracting party) generally cannot be set aside. If all the courts cared about was consent, the source of the mistake or pressure would be irrelevant.

Most contract scholars appear to regard the law dealing with incapacity, duress, undue influence and misrepresentation as core parts of the law of contract. With the occasional exception of incapacity, these topics are discussed in every contract law textbook, and they are introduced, so far as I am aware, without qualification or explanation. Some of the best-known 'contract law' articles focus on the law of duress.[39] There is a large literature devoted to the distributive consequences of contract law, nearly all of which focuses on duress, mistake and related doctrines.[40] The assumption underlying these articles is that in so far as duress, etc are based on distributive principles, contract law is based on distributive principles. More generally, nearly every discussion of what is widely assumed to be the core or most important contract law principle—the principle of freedom of contract—focuses on the above doctrines.[41] Thus the rules on capacity are presented as the law's attempt to ensure that parties have the general ability to make free choices, while the rules on

[38] See *ibid*, at ch 9.

[39] RE Barnett, 'A Consent Theory of Contract' (1986) 86 *Columbia Law Review* 269; JP Dawson, 'Economic Duress—An Essay in Perspective' (1947) 45 *Michigan Law Review* 253; J Dalzell, 'Duress by Economic Pressure' (1942) 20 *North Carolina Law Review* 341; RL Hale, 'Coercion and Distribution in a Supposedly Non-Coercive State' (1923) 38 *Political Science Quarterly* 470; AT Kronman, 'Contract Law and Distributive Justice' (1980) 89 *Yale Law Journal* 472.

[40] In addition to Dawson, Dalzell, Hale and Kronman, *ibid*, see, eg, PS Atiyah, 'Economic Duress and the Overborne Will' (1982) 98 *LQR* 197; J Gordley, 'Equality in Exchange' (1981) 69 *California Law Review* 1587.

[41] See, eg, M Trebilcock, *The Limits of Freedom of Contract* (Cambridge, Harvard University Press, 1993).

duress, undue influence and misrepresentation are presented as the law's attempt to ensure that parties have the ability to make free choices in particular cases.[42] Nonetheless, I suggest that these rules are not contract law rules, but belong instead to the general part of the law of obligations.

As already noted, contract scholars typically describe the basic event that gives rise to a contract as either a promise, a reliance-inducing statement, a transfer (of rights), or something similar. It is not a part of the ordinary description of any of these events that the participants have capacity, act free from duress, etc. A promise, for example, is created (merely) by intentionally communicating an intention to undertake an obligation.[43] Minors who understand the concept of a promise, as many do, are perfectly capable of making promises and most of them do so regularly. It is also perfectly possible to make a promise while under duress or while subject to undue influence or a misrepresentation. This is confirmed by ordinary language: we talk of 'promises made under duress', (not 'non-promises made under duress'). The distinction between the concepts of promise and duress also helps to explain why promises made under duress are sometimes binding. Every parent is familiar with the process of extracting promises from their children which are made under conditions that can only be described as duress. Yet the parents not only consider these statements to be promises, they consider them binding promises. The fact that Johnny was not allowed out of his room until he promised to return some toys to his brother does not excuse Johnny from keeping his promise. Every treaty signed by a surrendering army is signed under duress. In the common law, contracts that are made under duress are in general perfectly valid if the duress emanated from a third party. The same is true, in broad outline, of promises made while subject to another's influence or promises made as a result of a misrepresentation. This does not mean that promises made under duress, etc. should always or even often be legally binding. The point is merely that in so far as such promises are not binding, the reason is not that they are not promises but that, in the circumstances, the normal obligatory force of a promise is overridden or excused.

What is true of promises is equally true of reliance-inducing statements. It is clearly possible to induce someone to rely on your statements even if those statements are made under duress. The hostage who phones relatives to ask them to give the hostage-taker a sum of money is usually hoping that these relatives will rely on the call and in many cases they do just that.

[42] See, eg, C Fried, *Contract as Promise* (Cambridge, Harvard University Press, 1981).

[43] J Raz, 'Voluntary Obligations and Normative Powers' (1972) *Proceedings of the Aristotelian Society* Suppl 46; J Raz, 'Promises and Obligations' in PMS Hacker and J Raz (eds), *Law, Morality, and Society* (Oxford, Clarendon Press, 1977); Smith (2004), above n 6, at 57.

Few people would hold the hostage victim morally responsible for inducing such reliance, but this is not because there was nothing to fit the definition of reliance or because the reliance was unreasonable or unforeseeable by the one who asked for it. The question whether a transfer was made is similarly distinct from the question whether the transferor acted under duress, etc. A transfer happens whenever an individual intentionally transfers ownership to another. One can act intentionally even when one does not act freely. This is why transfers made under the threat of a severe sanction are sometimes perfectly valid. Many people pay taxes only because they are afraid of criminal charges, yet their payments are valid legally nonetheless. When a defendant transfers money to a claimant in compliance with a court order the transfer is valid even though the defendant had no real choice and the claimant was fully aware of this.

Admittedly, it remains possible, in theory anyway, that the basic contract-creating event is not merely a promise, reliance-inducing statement, transfer and so on, but something like a 'promise made by someone with capacity not acting under duress or subject to undue influence or a misrepresentation'. I noted earlier that at least some of the reasons for not enforcing a promise and so on cannot be explained by extrapolation from the idea of a promise, transfer and so on. The rule that promises made in domestic settings are assumed not to be legally binding is an example. Such promises are promises in the full sense, yet the rules regarding domestic promises are clearly a part of contract law. The law on duress, etc, might be explained in a similar way. For example, we might say that just as only non-domestic promises can create contracts, only 'freely made' promises, etc can create contracts.

The analogy breaks down, however, for two reasons. The first is that it makes sense—moral sense—to suppose that the rules on duress, etc are general rules. In broad terms, these rules express a principle of responsibility. The idea that one should not be held to contracts to which one did not consent reflects a notion of negative responsibility ('I am not morally responsible for my actions'), while the idea that one should not be able to enforce rights obtained through one's wrongdoing reflects a notion of positive responsibility ('I am responsible for your actions'). Responsibility in either sense is a complex notion.[44] Responsibility for a promise is different from responsibility for a transfer and each is different from responsibility for a wrong. But it seems highly unlikely that 'responsibility for' is relevant only for certain kinds of obligations and juridical acts or that the fundamental principles differ as between different obligations. The concept of responsibility is general—and so is outside the vertical border of contract law. This way of understanding responsibility has long been

[44] P Cane, *Responsibility in Law and Morality* (Portland, Hart Publishing, 2002).

adopted in the criminal law, where the distinction between the general and the special parts of the law is often explained primarily in terms of the distinction between responsibility and culpability. Not coincidentally, the various defences that make up the bulk of the general part of the criminal law (for example, duress, incapacity) are in kind (though not usually in degree) similar to the contract law defences of duress, incapacity, and so on.

The other reason the analogy between 'domestic promises' and 'promises made under duress' breaks down is that while the rules governing domestic promises are uniquely applied to contractual obligations, the rules on duress, etc are applied in non-contractual settings. Like the rules on damages and the rules on court orders, they are governed by general (that is, not specifically contractual) principles and they apply to a range of juridical acts. A promise to pay $100 that is made under duress (or undue influence, misrepresentation, etc) is not binding. Even assuming it is made in exchange for consideration and in a commercial setting and so on, it does not have the normal legal effect of a promise to pay $100. This is basic first year 'contract' law. Yet exactly the same conclusion is reached in cases where, instead of promising to pay $100, the $100 is transferred under duress by way of a gift, a trust or a bequest. Executed gifts, trusts, and wills are not contracts. Yet just as in the case of a (potential) contract, the normal legal effects of these events is in every case denied if the person initiating the transaction acted under duress. This is why the textbooks on unjust enrichment discuss incapacity, duress, undue influence and misrepresentation in detail. An unjust enrichment can arise because a contract was made under duress, but, equally, it can arise because a non-contractual juridical act (for example, a gift, a will, a trust) was made under duress. There is ancient authority that duress is no defence to an action in tort,[45] but few suppose the authority would be accepted today[46] (though there does not appear to be a modern decision on point). In any event, a defence based on duress can always be restyled as a defence based on necessity, which is accepted.

Admittedly, duress and other general defences do not operate identically across contractual and non-contractual settings. In part, the explanation for the differences is precisely that the relevant rules have traditionally not been understood as part of the general law. Contract cases in which these defences are raised are discussed and analysed separately from tort and unjust enrichment cases. This had led to inconsistencies. Indeed, there are inconsistencies in how these defences are applied within the law of contract. There has long been a recognised defence of presumed undue

[45] *Gilbert v Stone* (1647) Aleyn 35.

[46] WVH Rogers, *Winfield & Jolowicz on Tort*, 16th edn (London, Sweet & Maxwell, 2002) 876.

influence, but the courts have yet to recognise a defence of presumed duress, misrepresentation or incapacity. Fortunately, some scholars are beginning to examine these defences from a general perspective. Scholars working in the area of unjust enrichment law, in particular, have long been aware that concepts such as duress, undue influence and so on are important for both contract and unjust enrichment cases. More recently, scholars working on what might be called, roughly, the philosophical foundations of private law, have examined these concepts in general terms.[47] Consistent with the argument of this article, these scholars are attempting to do what criminal law scholars did for their subject long ago, namely, to present and analyse defences in general terms, abstracting from the particular offence or cause of action.

The main reason that the rules on duress etc appear to operate differently depending on the nature of the primary duty in question is, however, the same reason that the rules on damages and the rules on court orders appear to operate differently. The application of these rules to particular cases depends on the specific context, and the context of contract cases is typically (though not always) different from the context of unjust enrichment or tort cases. Just as the rules on remoteness cannot be reduced to a statistical test, the rules on duress, etc cannot be reduced to a mechanical test. As already noted, the defences of duress, incapacity and so on (like the rules of remoteness) reflect principles of responsibility. Responsibility is a complex, context-specific notion.[48] Responsibility for having made a promise is different from responsibility for a transfer and each is different from responsibility for a wrong. But the concept itself is general; it applies to contractual obligations as well as in other contexts. Any attempt to understand the rules that reflect this principle of responsibility must have regard to the full range of situations in which the principle is at issue.

VIII. CONCLUSION

In criminal law scholarship, the distinction between the special and the general parts of the law is well known. It has fostered the development of a sophisticated general criminal law of excuses, justifications and so on. The parallel distinction is almost unknown in the law of obligations, at least in common law jurisdictions. Part of the explanation for this situation is no doubt the usual lack of attention that, historically, the common law has

[47] Cane, above n 7; W Lucy, *Philosophy of Private Law* (New York, Oxford University Press, 2007) 47–144; P Birks, 'The Concept of a Civil Wrong' in D Owen (ed), *Philosophical Foundations of Tort Law* (Oxford, Clarendon Press, 1995) 29.

[48] Cane, *ibid*.

paid to classificatory undertakings generally. A more explicit reason is that the practical consequences of neglecting to distinguish between the general and the specific are less evident and certainly less immediate than those arising from neglecting the distinction between the different special parts of obligations law. But there are consequences. Failing to pay attention to contract law's vertical border can lead scholars, judges and lawyers to apply the wrong principles to understand legal rules, to draw the wrong inferences from those rules, and to fail to make appropriate generalisations. The frequent attempts by contract law scholars to draw conclusions about the nature of contracts from the rules governing damages and specific performance is just the most obvious example of where this has happened.

This article has barely scratched the surface. There is a significant amount of law that was not discussed (for example, formalities, unconscionability, unenforceable contracts) and what law was discussed was examined only in the briefest outline. The article's substantive conclusions are only provisional. The article's primary aim, however, was not to pronounce final judgment on the limits of contract. Rather the aim has been to introduce the idea that contract law has limits, to explain why these limits are important, and to give some idea of how they should be identified.

2

Border Control: Some Comparative Remarks on the Cartography of Obligations

HELGE DEDEK*

I. INTRODUCTION

A. Mapping Law's Empire

LEGAL HISTORY HAS witnessed many attempts to impose a measure of order onto the unpruned growth of the common law, whether it be through codification, restatement or scholarly treatment. It has particularly been the ambition of jurists, at least since William Blackstone's time, to provide some guidance to those who, taking recourse to the law, are bound to get lost in its wilderness. Legal scholars strove for exploration and intellectual conquest. 'Mapping' is one of the most frequently recurring metaphors for describing those scholars' endeavour—mirroring the enterprise of their counterparts, adventurers and scientists exploring[1] and measuring[2] the real world instead of law's virtual reality.[3] Underlying these projects is what Pierre Schlag has called the 'legal aesthetic of the grid', which

* I would like to thank Richard Bronaugh, Nicholas Kasirer, Stephen Smith and Stephen Waddams for their support and helpful comments. I also would like to thank Jason Neyers for his kind invitation.

[1] See, for an introduction, D Arnold, *The Age of Discovery*, 2nd edn (London, Routledge, 2002) *passim*. See also DJ Boorstin, *The Discoverers* (New York, Random House, 1983), and P Whitfield, *New found lands: Maps in the History of Exploration* (London, Routledge, 1998), who presents the intellectual dimension of (real-world) exploration through the history of (actual) maps and the science of cartography.

[2] For a colourful literary account of the scientific spirit of appropriating the unknown by exploring and measuring see D Kehlmann, *Measuring the World*, trans CB Janeway (London, Quercus, 2007), who focuses on the travels of obsessive-compulsive nineteenth century explorer–scholar Alexander von Humboldt.

[3] S Waddams, *Dimensions of Private Law* (Cambridge, Cambridge University Press, 2003) 1ff; G Samuel, 'Can the Common Law Be Mapped?' (2005) 55 *University of Toronto Law Journal* 271. For other 'legal science' metaphors invoking taxonomic disciplines such as

frames 'the law as a field, a territory, a two-dimensional space that can be mapped and charted'.[4] Legal concepts are assigned a spot on the map, so that, eventually, everything, as Christopher Columbus Langdell wrote, 'should be found in its proper place, and nowhere else'.[5]

A debate has been revolving for centuries around the merits of 'mapping' the law, with voices of criticism particularly and persistently emanating from the bench.[6] Its most recent revival, the ensuing of a new 'taxonomy debate',[7] is, of course, owed to the works of the late Peter Birks of Oxford. Birks, in what is probably the most well-known articulation of his ambitions, had once more adopted the geographical imagery and the metaphor of the map. After all these years, writes Birks, lawyers are still lost and fall victim to what he calls a 'stovepipe mentality'—knowing 'their law only in the way that many people know London, as pools of unconnected light into which to emerge from a limited number of friendly tube stations'.[8] The reason for that mentality is, according to Birks, that 'nobody has shown them the *map*'.[9]

What is new, what is so remarkable about this last attempt at 'mapping'? Maybe, and tentatively, a twofold answer can be given to this question: first, because it is so openly and courageously anachronistic. At the turn of the century, after the worldwide impact of Legal Realism and the statements of many scholars who wrote off the 'mapping' enterprise as a remnant of nineteenth century legal thought,[10] this move has come rather unexpectedly; and it is not surprising that it originated in England where

anatomy, see, eg, AJ Rodenbeck, *The Anatomy of the Law, A Logical Presentation of the Parts of the Body of the Law* (Boston, Little, Brown & Co, 1925) and the caustic review by E Munguia Jr, (1925–26) 14 *California Law Review* 150–52. On the role of metaphors for change and progress in scholarship in general see JG Daugman, 'Brain Metaphors and Brain Theory' in W Bechtel *et al* (eds*), Philosophy and the Neurosciences* (Oxford, Blackwell, 2001) 23–5; TS Kuhn, 'Metaphor in Science' in A Ortony (ed), *Metaphor and Thought* (Cambridge, Cambridge University Press, 1993) 409–19. See also N Kasirer, 'Bijuralism in Law's Empire and in Law's Cosmos' (2002) 52 *Journal of Legal Education* 29, 32, 34 *et passim*, who deliberately uses the imagery of the 'map' to characterise a jurisdictional, territorial way of thinking about law as opposed to a more abstract, 'borderless' approach.

[4] P Schlag, 'The Aesthetics of American Law' (2002) 115 *Harvard Law Review* 1047, 1055.

[5] CC Langdell, *A Selection of Cases on the Law of Contracts*, 2nd edn (Boston, Little, Brown & Co, 1879) ix.

[6] See, eg, the more recent statements in *Read v J Lyons & Co* [1947] AC 156 (HL) 175; *Woodar Investment Development Ltd v Wimpey Construction UK Ltd* [1980] 1 WLR 277 (HL) 283; *Attorney General v Blake* [2001] 1 AC 268 (HL) 284 [*Blake*].

[7] G Samuel, 'English Private Law: Old and New Thinking in the Taxonomy Debate' (2004) 24 *OJLS* 335.

[8] P Birks (ed), *English Private Law*, 1st edn (Oxford, Oxford University Press) vol 1, xxxv.

[9] *Ibid,* at xxxvi (emphasis added).

[10] Just see Schlag, above n 4, at 1055ff. See also D Kennedy, 'Toward an Historical Understanding of Legal Consciousness: The Case of Classical Legal Thought in America, 1850–1940' (1980) 3 *Research in Law & Sociology* 3.

the ideological impact of the Realists (and the following deconstructive movements that built on their legacy) was never as strong as in the New World.[11] However, what seems more, and, in the true sense of the word, anachronistic, is that Birks, a Romanist, openly takes recourse to, as Andrew Burrows puts it, 'his beloved Gaius',[12] namely, the institutional scheme ascribed to classical Roman jurist Gaius, who lived and taught law in the second century AD.[13] Gaius's Institutes were to Birks the paradigm of scholarship successfully 'providing a map'; he sees the English lawyers' lack of orientation as a result of the lack of exposure to the Institutes in modern legal education.[14] This inspiration for the classificatory scheme employed by Birks, and Birks's openness about the source of his inspiration, leads us to the possible second part of our answer: the Birksian call for taxonomy is also remarkable because it seems so blatantly *civilian*.

Civilian scholarship has pushed its obsession with taxonomy and classification much further than the common law tradition; the 'formal rationality' (Max Weber) of the civil law is one of the stereotypes that emerge whenever attempts are made to describe the differences between the common and the civil law.[15] Particularly to those who see legal discourse as a unique and peculiar expression of 'legal culture', imposing a civilian classificatory scheme—*the* most famous civilian classificatory scheme—onto the common law must appear to be a problematic project due to the incompatibility of 'mentalities'.[16] Bringing the English common law closer to the Continent by imbuing it with a civilian spirit of order, the Birksian project has political implications as well, particularly against the backdrop of the ongoing efforts to create a 'European Private Law' employing the civilian method of codification.[17]

The merits of taxonomy and classification in law have, of course, been discussed by many on a more abstract and general level.[18] A 'classical'

[11] See, eg, N Duxbury, 'English Jurisprudence between Austin and Hart' (2005) 91 *Virginia Law Review* 1.

[12] A Burrows (ed), *English Private Law*, 2nd edn (Oxford, Oxford University Press, 2007) xxxi.

[13] See, eg, HLW Nelson, *Überlieferung, Aufbau und Stil von Gai Institutiones* (Leiden, Brill, 1981) *passim*. On 'Gaianism' see also DR Kelley, 'Gaius Noster: Substructures of Western Social Thought' (1979) 84 *The American Historical Review* 619ff.

[14] P Birks, 'Definition and Division, A Meditation on Institutes 3.13' in P Birks (ed), *The Classification of Obligations* (Oxford, Clarendon Press, 1997) 1ff.

[15] See, eg, JH Merryman, *The Civil Law Tradition*, 2nd edn (Stanford, Stanford University Press, 1984) 90ff.

[16] See P Legrand, 'European Legal Systems Are Not Converging' (1996) 45 *Intellectional and Commercial Law Quarterly* 52, 60ff; P Legrand, 'How to Compare Now?' (1996) 16 *Legal Studies* 232, 237ff.

[17] Which, in itself, gives rise to hostility from the proponents of 'law & culture': P Legrand, 'Antivonbar' (2006) 1 *Journal of Comparative Law* 13.

[18] See, eg, Schlag, above n 4. On the impossibility of 'legal science'—likening it to the pseudo-science of 'phrenology'—see P Schlag, 'Law and Phrenology' (1997) 110 *Harvard Law Review* 877.

treatment of the subject is Roscoe Pound's article on 'classification of law', disenchanting the idea that a 'formal logic' might operate in legal taxonomies.[19] Stephen Waddams has also expressed scepticism towards the 'concept of legal mapping' in his study *The Dimensions of Private Law*, drawing on an in-depth analysis of common law reasoning.[20] The works of Geoffrey Samuel stand out as a direct response to the works of Birks, critically scrutinising, for example, the value of Birks's statement that 'Gaius was the Darwin of the law'.[21]

However, it is not my intention, in this venue, to partake in the general theoretical discussion around the raison d'être of legal taxonomies. This is not the place for an inquiry that would involve philosophical, psychological, anthropological, sociological and linguistic questions. Rather, I want to approach legal taxonomy from a perspective that might be called internal, looking at some of the effects and ramifications of legal categorisation within legal discourse.[22] I shall, therefore, work from (overly) simple axioms, by treating the answer to the following two questions as settled. First (starting from Samuel's object of scrutiny): Is Gaius for the law what Darwin was for science? Can, in other words, law be likened to zoology or any other natural science? I simply assume that the answer is: No, obviously not. Of course, the philosophy of science has brought us the insight that neither scientific taxonomies[23] nor even maps[24] are just representations of the real world, but always a construction as well. Scientific taxonomies are not different from legal ones simply because the

[19] R Pound, 'Classification of Law' (1924) 37 *Harvard Law Review* 933–69.

[20] Waddams, above n 3, at 222ff.

[21] G Samuel, 'Can Gaius Really Be Compared to Darwin?' (2000) 49 *ICLQ* 297. See also Samuel, above n 7, and Samuel above n 3.

[22] For the same reason, I shall not engage in an analysis of the methodological steps (conceptualisation, categorisation, classification) involved in building a taxonomical 'system' and shall not differentiate between terms such as 'categorisation' and 'classification' with the technical denotations assigned by other disciplines. See, for such definitions, R Ellen, 'Introduction: Categories, Classification and Cognitive Anthropology' in R Ellen (ed), *The Categorical Impulse, Essays in the Anthropology of Classifying Behaviour* (New York, Berghahn, 2006) 1.

[23] This idea is usually illustrated by referring to Michel Foucault's comment (*Les mots et les choses: Une archeologie des sciences humaines* (Paris, Gallimard, 1966) 7) on Borges's apocryphal 'certain Chinese encyclopedia': 'Dans l'émerveillement de cette taxinomie, ce qu'on rejoint d'un bond, ce qui, à la faveur de l'apologue, nous est indiqué comme la charme exotique d'une autre pensée, c'est la limite de la nôtre: l'impossibilité nue de penser cela.' For an English translation, see M Foucault, *The Order of Things* (New York, Random House, 1973) xv. The understanding of taxonomies as relative to culture rather than correct or incorrect depictions of reality makes them a favourite subject of anthropologists and linguists; see, eg, Ellen, *ibid* and the study by G Lakoff, *Women, Fire, and Dangerous Things, What Categories Reveal about the Mind* (Chicago, University of Chicago Press, 1990). On the other hand, cognitive psychology has insisted that categorisation of 'the real world' does not happen completely arbitrarily; see, eg, CB Mervis and E Rosch, 'Categorisation of Natural Objects' (1981) 32 *Annual Review of Psychology* 89, 91ff.

[24] See, eg, JB Harley, 'Deconstructing the Map' (1989) 26 *Cartographica* 1–20.

former depict things that physically exist, whereas the latter order abstract concepts that only exist in our minds. However, the categorisation of legal rules is special due to its normative implications. Categorising a rule, for example, as 'contractual' rather than 'tortious' invokes a plethora of denotations[25] that characterise, consciously and subconsciously, the category of 'contract': exercise of free will and consent as opposed to involuntary injury, distributive rather than corrective justice and so on. These properties of the category, in turn, inevitably influence the way a particular rule—once placed in this category—is understood, construed and justified. Categorisation and classification in the law—and this is a thought we will return to later—is not just descriptive, but prescriptive as well. We cannot treat legal concepts as entities that can be 'objectively' classified by family resemblance, needing simply to be spiked on the proper pin to be displayed in a cabinet.

Given this answer, is it useless or even, as some scholars seem to think, *ridiculous* to work with the idea of a taxonomy, a classification of legal concepts?[26] Again, the answer is: No, obviously not. Categorisation and classification are not restricted to 'taxonomic' disciplines such as zoology. Rather, they are the basic cognitive processes that permit the brain to reduce the complexity of stimuli to a degree that it can make sense of the world.[27] These processes are the precondition of the apprehension of the natural world as well as the world of our own thoughts. Without categorisation and classification, there would be no language to talk about law in the first place; without *some* kind of taxonomy, lawyers could not communicate, nor could laws be compared. Most importantly, taxonomies are indispensable to legal education; let us not forget that Gaius's as well as Blackstone's mapping projects were driven by their pedagogical ambitions: 'A *plan*[28] of this nature,' writes Blackstone, 'if executed with care and ability, cannot fail of administering a most useful and rational entertainment to students of all ranks and professions.'[29] Once we understand and acknowledge the *instrumental* rather than 'natural' character of legal categories and classifications, there is nothing wrong with developing categories and classifications that soundly serve their purpose. Roscoe Pound summarised it thus:

[25] The idea that a multitude of information—defined by 'communal traditions and communal processes'—finds expression in certain 'paradigms' that guide legal classification is also expressed by JM Feinman, 'The Jurisprudence of Classification' (1989) 41 *Stanford Law Review* 661, 699.

[26] D Campbell, 'Classification and the Crisis of the Common Law,' book review of *The Classification of Obligations* by P Birks (1999) 26 *Journal of Law & Society* 369, 370: 'perfectly laughable'.

[27] See Lakoff, above n 23, at 5ff; Ellen, above n 22, at 1ff.

[28] Emphasis in the original.

[29] W Blackstone, *Commentaries on the Laws of England*, facsimile of 1st edn 1765–1769 (Chicago, University of Chicago Press, 1979) vol 1, 36.

> *Classification is a shaping and developing of traditional systematic conceptions and traditional systematic categories in order to organize the body of legal precepts so that they may be: (1) Stated effectively with a minimum of repetition, overlapping, and potential conflict, (2) administered effectively, (3) taught effectively, and (4) developed effectively for new situations.*[30]

B. Demarcating Contract Law's Territory

The title of this volume is 'Exploring Contract Law'; given the historical affinity of 'exploration' and 'mapping,' it is not surprising that also in this volume we should find a contribution that engages in the enterprise of 'mapping' contract law.[31] The activities of 'exploring' and 'mapping' are not limited to *terra incognita*. Indeed, Stephen Smith makes a compelling argument that some areas that are commonly considered to belong to 'contract law' are not specifically 'contractual' at all. Rather, writes Smith, perhaps certain domains of contract law could be seen as belonging to a general part of the law of obligations. While we thought we knew the map—maybe not, to speak with Blackstone, down to every inconsiderable hamlet[32] (this very fact, however, is what keeps the cartographers going!)—Smith redefines the whole territory, redraws the borders and lines of demarcation, and proclaims independence for an area hitherto unrecognised by the common law.

As I said, I do not want to engage with the question of whether this enterprise is a legitimate or a useful one. Let us just assume that it is, and let for now a simple justification suffice, namely, that thinking about a certain doctrine in a different way—for example, changing its classification—may shed new light on the purposes it serves. At the very least, it might help us to question the traditional explanations and justifications we are taught to believe in. This critical inquiry, therefore, is a process that has an intellectual value in itself.[33] In this article, however, I want to point to some of the problems that a classification, in particular a division into the 'general' and the 'special' might entail. I would like to demonstrate how a classification that excludes certain areas from contract law as such runs the risk of obscuring certain connections that can only be seen when looking at contract law in a more inclusive way.

I want to do so by assuming a comparative perspective and reporting on the experience of German civil law. Not being familiar with a comprehensive category of contract law, German civil law has, as a consequence,

[30] Pound, above n 19, at 944 (emphasis in original).
[31] SA Smith, 'The Limits of Contract' ch 1.
[32] Blackstone, above n 29, at 35.
[33] The procedural character, the 'process nature' of classification is also mentioned by Feinman, above n 25, at 704.

encountered certain doctrinal difficulties, particularly in justifying the award of what is usually called 'reliance damages,' which a less fractioned view could have helped to avoid. This thought will lead us to the more fundamental problems associated with taxonomising the law, namely, the danger of the proverbial 'hardening of categories,' that 'well-known ailment of lawyers'.[34]

II. DIFFERENT TRADITIONS, DIFFERENT MAPS: THE 'GENERAL' AND THE 'SPECIAL'

The move for a taxonomic ordering of the common law has at all times been to some extent inspired by civilian ideas. As Smith acknowledges, the cartography he suggests displays, just like the Birksian writings on taxonomy, a certain affinity to civilian ideas.[35] In some sense, Smith goes even further: his proposition is *beyond* just civilian—it is almost *German*. The idea that contract law should not only be subdivided but also curtailed in a way that makes room for a 'general part of obligations' and even a 'general part' of private law as such is a particularly German–civilian categorisation. The taxonomists of the common law have also long been familiar with the technique of classifying into 'general' and 'special,' the most prominent of these taxonomists being, of course, Jeremy Bentham.[36] However, many common law writers who have tried to make use of this theme to guide their taxonomic efforts have explicitly looked at the classificatory structure of the German civil law.[37] This structure deviates from the classical Gaian institutional scheme and from the organisation of the Code Napoléon as well.

A. Civil Law v Civil Law: The Hallmarks of German 'Legal Science'

By 'not only civilian, but German' I also mean that certain hallmarks of civilianism, in particular the tendency to formalism and taxonomisation,

[34] Lord Nicholls in *Blake*, above n 6, at 264 (HL) 284, quoting JP Dawson, 'Restitution or Damages' (1959) 20 *Ohio State Law Journal* 175, 187.

[35] Smith, above n 31 at fn 1.

[36] J Bentham, 'Papers Relative to Codification and Public Instruction: Including Correspondence with the Russian Emperor, and Divers Constituted Authorities in the American United States' in P Schofield and J Harris (eds), *'Legislator of the World': Writings on Codification, Law, and Education* (Oxford, Clarendon Press, 1998) 8.

[37] See, eg, A Kocourek, 'Classification of Law' (1934) 11 *New York University Law Quarterly Review* 319, 327: 'The German Code has become the model of classification of our time.' After sketching a taxonomy of taxonomies, Kocourek attempts to develop his own 'scientific classification,' involving an elaboration of a 'Theory of a General Part,' comparing his ideas to the actual organisation of the German code: *ibid*, at 341ff.

have been pushed to an extreme in the German civil law tradition, much further than in France. The differences among the various expressions of the civilian tradition must not be underestimated. This is, of course, a trite statement, but given how many stereotypes circulate about 'the' civil law, it is sometimes helpful to remember that the respective national civilian systems are not alike in the way that common law systems are alike in the Commonwealth. Civilian systems may share the common past of the Roman law-based *ius commune* and, at least on the Continent, the experience of codification. However, we must not forget that codification itself ingrained a positivistic solipsism which had the consequence of almost completely cutting off discursive exchange between national legal systems on the Continent. Furthermore, the respective codes are already the product of different philosophies, styles, fashions and of the growing importance of the process of 'nationalising' the law: the 'age of codification' was the same age that, on the Continent, witnessed the birth of the modern nation-state.[38]

If we compare the French and the German codes and try to describe their distinctive features, we are immediately confronted again with the perils of classification, this time in a different context: history (and the social sciences in general) has to work with conceptual classifications such as 'epochs' and 'ideas'. Not unlike what the jurist tries to do by conceptually contextualising doctrines, rules and principles, the historian tries to explain past events by fitting them into a broader context and by grouping them together under certain labels.[39] The result is, again, ambivalent. Epochs and periods might indeed have something like a certain spirit that was in a way characteristic and unique, and pointing out this specific property helps one better understand historical events. However, by branding a certain era with a certain mark we also construct an oversimplified identity and difference that does injustice to continuities and similarities.

Keeping this in mind, we might say that the French Code Napoléon grew out of the eighteenth century spirit of Natural Law, expounding broad principles and making fundamental statements about justice and citizens' rights; whereas the German Code, a century younger, breathes the technocratic formalism of the nineteenth century, an age that endowed the idea of 'legal science' (with all its implications for legal taxonomy and classification)

[38] On the nationalisation of the *ius commune* (the basis of the codes!) by nationalising *institutional* writing see K Luig, 'The Institutes of National Law in the Seventeenth and Eighteenth Centuries' (1972) 17 *Juridical Review (NS)* 193, 195: 'This new law was a complete and coherent unit, and theoretically independent of the earlier European *ius commune*; it was the expression of the national independence and autonomy of the rising nation-states of the *ancien régime*.'

[39] See, eg, C Behan McCullagh, 'Colligation and Classification in History' (1978) 17 *History and Theory* 267ff.

with a whole new meaning.[40] Of course, this would be an over-simplification. The 'scientific' approach to law and the development of taxonomic schemes was not an invention of the nineteenth century; for the longstanding nature of this 'scientific' tradition, we might point to Leipniz, the *mos geometricus* championed by the enlightenment in general,[41] to Petrus Ramus,[42] or we might go back to the tables of distinctions of the medieval scholastics[43] and finally, of course, to Gaius himself. However, the approach to law *did* change again in an age that saw industrialisation and the rise of the market and believed in progress through the advancement of science and technology. Savigny's Historical School, which proclaimed that the *ius commune* could be improved by an informed study of the classical Roman sources, employing state-of-the-art methodologies of linguistics and history, turned slowly into 'the science of the Pandects'[44] whose ideal of the law was that of an immaculate clockwork. Cutting-edge scholarship had, if you will, given way to engineering. High-end engineering, that is: centuries of law as an academic discipline and an ideal of intellectual rigour befitting the professor (just recall Max Weber's 'Science as a Vocation') drove German nineteenth century legal scholars to push their conceptual jurisprudence to an unparalleled degree of mechanical precision and technical perfection. While again walking the thin line between attaching a helpful label and falling victim to cliché, we might say that this expresses, better than anything else, the 'Germanness' of German private law—a veritable field day for those who see connections between legal culture and national character.[45] However, if you are interested in how a jurisprudential style that is safely grounded in a highly developed 'scientific' taxonomy operates, looking at German law is definitely the right thing to do.

[40] See, eg, K Zweigert and H Kötz, *An Introduction to Comparative Law*, 3rd edn, trans T Weir (Oxford, Clarendon Press, 1998) 145.

[41] See, eg, R Berkowitz, *The Gift of Science* (Cambridge, Harvard University Press, 2005) 17ff.

[42] JS Freedman, 'The Diffusion of the Writings of Petrus Ramus in Central Europe, *c* 1570–*c* 1630' (1993) 46 *Renaissance Quarterly* 98ff. See also DR Kelley, 'Jurisconsultus Perfectus: The Lawyer as Renaissance Man' (1988) 51 *Journal of the Warburg and Courtauld Institutes* 84, 91ff.

[43] G Otte, *Dialektik und Jurisprudenz, Untersuchungen zur Methode der Glossatoren* (Frankfurt am Main, Vittorio Klostermann, 1971) *passim*.

[44] F Wieacker, *Privatrechtsgeschichte der Neuzeit*, 2nd edn (Göttingen, Vandenhoeck & Ruprecht, 1967) 430.

[45] See, eg, OG Chase, 'Legal Processes and National Culture' (1997) 5 *Cardozo Journal of International & Comparative Law* 1 (and see of course JH Langbein's angry reply 'Cultural Chauvinism in Comparative Law' (1997) 5 *Cardozo Journal of International & Comparative Law* 41). For a far more subtle elaboration on the 'Frenchness' of French civil law see Legrand, above n 16, at 235ff; N Kasirer, 'Larger than Life' (1995) 10 *Canadian Journal of Law and Society* 185. National peculiarities were of course tempered by actual intellectual exchange: the 'Pandectistic' style of jurisprudence was, in the nineteenth century, of great allure for all civilians (and, incidentally, for many common lawyers as well!) and of major influence in jurisdictions that had adopted 'Natural Law Codifications' earlier, such as Prussia and Austria, and even France: French civil law scholarship was 'pandectised'. Let us not forget that the canonical Aubry/Rau started out as a Pandectist treatise.

Whether one sees a pinnacle of plain orderly elegance or a pedantic nightmare of anachronistic formalism is in the eye of the beholder alone.

B. Stephen Smith's 'Limits' vis-à-vis the Limited German Concept of 'Contract Law'

(i) Horizontal and Vertical Borders

Having refreshed this bit of background information, we can turn back to contract law. In his contribution to this volume, Stephen Smith distinguishes two sets of borders that demarcate contract law's territory. First, there is a 'horizontal' border, which 'separates contract law rules from the other sets of primary obligation-specific rules that occupy the bottom of the obligations pyramid'[46] as, for example, the obligation not to trespass, the obligation not to damage another's property, the obligation not to defame and so on. Secondly, Smith defines the 'vertical' border of contract law as the one that

> sets the boundary between contract law and the general part of the law of obligations. This border is vertical because it sets the boundary between rules at the top and bottom of the pyramid; specifically it distinguishes contract law rules (and other primary-obligation rules) from rules that are located above them in the sense that they qualify these rules or regulate the consequences of breaching the obligations to which these rules give rise.[47]

To find out whether a rule 'vertically' belongs to the law of contracts, Smith asks two questions: '(1) is the rule applied to resolve not just contractual but also non-contractual disputes? (2) is the rule of a kind that it makes sense, morally, to suppose should be applied to obligations generally as opposed merely to contractual obligations?'[48] Asking the latter question implies an assumption we briefly mentioned before, namely, that taxonomy in law is never only descriptive, but also prescriptive in nature.

(ii) The 'Law of Obligations' in the German Civil Code

Smith acknowledges explicitly that his call for 'vertical borders' bears a certain resemblance with civilian ideas and, in particular, the German way of defining the scope of contract law in respect to a 'general part' of the 'law of obligations'.[49] Indeed, the German Civil Code is built around the principles that repetition should be avoided, and that rules that are to be

[46] Smith, above n 31, in the text accompanying fn 4.
[47] *Ibid*, text accompanying n 5.
[48] *Ibid*, in the text at end of Section III.
[49] *Ibid*, in the text accompanying n 1.

applied generally and not just in one particular context are to be systematically grouped together to form a 'general part'.[50] Thus, the codal chapter ('book') on the law of obligations is divided into subsections, whose ambit moves from the rules applicable to all obligations, to those rules applicable to all contractual obligations, and then, finally, to the rules only applicable to 'specific obligations.' Only the eighth subsection deals with specific obligations: it is subdivided into 27 'titles' which first deal with specific contracts (sale, loan, donation and so on) and then, in the last two titles, with unjustified enrichment and delicts. Rules that are supposed to be even more general in the sense that they are to be applied not only in the law of obligations but also to all other parts of the Code (the 'law of things', family law, hereditary law) are clustered together in the 'General Part,' §§ 1–240, as are, for example, the rules on personhood, age of consent and so on.

(iii) Breaking Away from Gaius: Origins of the 'Pandektensystem'

This regulatory style, moving from the general to the specific, culminating in the overarching abstractness of the 'General Part', has been seen as the unique characteristic of the German civil law.[51] The underlying technique has been likened to the mathematical operation of 'factoring out', isolating a factor in an equation and placing it in front of parentheses.[52] As expounded by Andreas Schwarz in the still canonical treatment of the subject,[53] the historical background of this 'Pandektensystem' and its particular 'mathematical' approach is the osmosis of Natural Law ideas regarding the systematisation of law into the scholarly treatment of the *ius commune*. Natural lawyers Pufendorf and then, most notably, Christian Wolf, developed 'logical' legal systems in the abstract, which, although without any significance for the law in action, inspired *ius commune*

[50] For a concise introduction to the 'regulatory technique' of the BGB see BS Markesinis, *The German Law of Contract,* 2nd edn (Oxford, Hart Publishing, 2006) 16ff. For an introduction to the historical reasons for this particular approach, see R Zimmermann, *The Law of Obligations* (Oxford, Oxford University Press, 1996) 29ff. See also Kocourek, above n 37, at 327.

[51] See, eg, F Pollock, *Essays in Jurisprudence and Ethics* (London, Macmillan, 1882) 14 (looking at Pandectist writing). This technique has, taking part in the process of 'Pandectisation' on the Continent (see above n 45), influenced other civilian traditions as well, even if the general–special dichotomy is not reflected in their respective codes. Pound, therefore, calls the arrangement that singles out a 'general' part that of 'the modern Civilian' (Pound, above n 19, at 947, 962ff).

[52] This common metaphor seems to have been first employed by G Boehmer, *Grundlagen der Bürgerlichen Rechtsordnung,* (Tübingen, JCB Mohr (Paul Siebeck), 1951) vol 2, pt 1, 72.

[53] AB Schwarz, 'Zur Entstehung des modernen Pandektensystems' (1921) 42 *Zeitschrift der Savigny-Stiftung für Rechtsgeschichte Romanistische Abteilung* 578.

scholars to deviate from the traditional institutional scheme.[54] Introduced—but later abandoned again—by Gustav Hugo in his treatise 'Institutionen des heutigen römischen Rechts' (1789), it was fully developed by Georg Arnold Heise in his work 'Grundriß eines Systems des gemeinen Civilrechts zum Behuf von Pandecten-Vorlesungen' (1807). However, we do not know whether this 'Pandektensystem' would have been immortalised in the structure of the Civil Code a century later, had it not been for the endorsement of the most prominent heads of German private law scholarship,[55] among them Friedrich Carl von Savigny himself, the towering figure of German nineteenth century legal scholarship. Savigny adopted Heise's system for his lectures and thus helped to make it the standard order in which the structure of the German *ius commune* was understood.[56] This perception then determined the structure of the Code; to this very day, it provides the outline that defines the 'classical' scope of treatises on private law and, most importantly, the curriculum of private law education.[57] General Part, General Part of the Law of Obligations, Specific Part of the Law of Obligations (divided into contractual/non-contractual obligations) is still the typical *cursus* law students have to pass through. The taxonomy, thus, moulds and dominates the mindset of each and every lawyer.

(iv) The Interplay of the General and the Special: Examples and Problems

Let us try to understand the effects of this compartmentalised approach by looking, for reasons of convenience, at some of the examples treated by Smith. We already briefly addressed the rules on personality. These rules apply not only in the context of contractual obligations, but determine, according to the logic of the Bürgerliches Gesetzbuch (BGB), whether a person—legal or 'natural'—can be the subject of any kind of right, be it in the context of contract, property or family law. Therefore, according to the

[54] *Ibid*, at 583. French law, of course, has seen similar developments, but with different consequences for the code. See, most notably, Domat's *Les droits civiles dans leur ordre naturel* (1689–97).

[55] Schwarz, above n 53, at 580.

[56] M Avenarius, 'Der Allgemeine Teil des Obligationenrechts aus Savignys Pandektenmanuskript-Bedeutung und Grundsatze der Edition' in M Avenarius (ed), *Friedrich Carl v Savigny, Pandekten, Obligationenrecht, Allgemeiner Teil*, (Frankfurt am Main, Vittorio Klostermann, 2008) 12ff.

[57] It is interesting that Pound (above n 19, at 940) remarked disapprovingly on the fact that, vice versa, the organisation of the Code had been influenced by the pedagogically motivated arrangement of Windscheid's famous textbook on the Pandects, which, according to Pound, might not have been the optimal arrangement for a restatement of the law. However, properly understood, Windscheid's treatise was nothing less than a restatement of the law in itself.

taxonomy of the BGB, these rules are located not in the general part of the law of obligations, but in the 'General Part' of the whole Code.

(a) Formation of Contract There, in the 'General Part,' we also find the rules on mistake, duress, unconscionability and illegality. This is surprising—why would those doctrines be treated outside the law of obligations? Their placement is owed to the fact that the rules on 'legal transactions' (Rechtsgeschäft) and 'declarations of will' (Willenserklärung) are located here, which actually means that the rules on offer and acceptance and the 'meeting of the minds' are not dealt with as part of the law of obligations at all. This, of course, is even more surprising—are not those rules, of all rules, the most 'contractual' ones? Yes, they are. But in aspiring to distill the general ideas out of specific rules and principles, German private law developed a general concept of 'contract' beyond just contracts that create obligations. This, of course, must seem utterly enigmatic to the common lawyer who is not even able to see a 'gift' as a (donative) contract because it does not entail promises of future performance. The most prominent example of the non-obligatory contract would be an agreement that transfers ownership.[58] In German law, an obligatory contract, such as a contract of sale, does not per se transfer ownership. The transfer of ownership requires the transfer of (corporeal) possession (*traditio* or an equivalent), and, in addition, *another separate agreement in which the parties agree on the transfer of ownership*. This second agreement does not create an obligation, but alters legal *status*—in our example, the allocation of ownership. This agreement would still be seen as a 'contract.' With contract thus defined, it makes sense (if only according to the twisted logic of the unique understanding of 'contract' in German law) to place the rules on formation of contract in the parts of the Code that are supposed to be applicable not just to contract law, not just to obligations, but to all agreements in private law, be they the source of obligations or not.

In this specific institutional context, the first part of Smith's twofold test of whether a rule should be seen as 'general' or not would be satisfied. What about the second part of Smith's test: does it make good *moral* sense to treat it as a general rule? This is obviously a more complicated question. If we had the chance to ask the historical *spiritus rectores* of this

[58] See, on the so-called 'Principle of Separation' and the 'Principle of Abstraction' that define the 'relationship of obligation' and the 'transaction that transfers, alters, [etc], rights' as separated and independent of each other, Markesinis, above n 50, at 27. The broad concept of 'contract' beyond agreements giving rise to obligations was developed by Friedrich Carl von Savigny, *System des heutigen römischen Rechts*, Bd III (Berlin, Veit und Comp, 1840) § 140, 312ff. On the significance of this development for the placement of 'legal transactions' in the 'General Part' see JF Stagl, 'Die Rezeption der Lehre vom Rechtsgeschäft in Österreich durch Joseph Unger' [2007] *Zeitschrift für Europäisches Privatrecht* 37, 39.

taxonomy, some might be confused by the implication that the functioning of the clockwork would have anything to do with morality.[59] It is in this sense that Smith transcends his predecessors in the taxonomic enterprise, who would have described their project as exclusively involving logics, not morals.

However, others, and among them the titanic Savigny himself, would have answered, I am sure, in the affirmative as follows: the capacity of a person to bring obligations into life and to create, transfer and dispose of rights by their sheer willpower is more than contract law mechanics, it is an expression of the autonomy of the will as such. The autonomous will is a condensation of the idea of individual freedom, which is the central moral concept in early nineteenth century German private law.[60] Private law demarcates spheres of individual freedom in a way that 'assigns the individual will a region where it can reign unperturbed by any other will'.[61] According to this ideology, the technical rules which are the most immediate emanation of this principle deserve to be exposed in the 'General Part' as a statement about the fundamental significance of the will in private law. Having crystallised in the structure of the Code, this ideology is also set in stone and remains so in a changing intellectual environment.[62] The question suggested by Smith has to be asked again and again to prevent the 'hardening of categories' into anachronistic fossils.

(b) Contract Damages The codal classification of the rules on damages depends, likewise, on the degree of their generality. Being applicable only to obligations, *sedes materiae* of those rules is not the 'General Part'. However, rules that govern contractual as well as tortious obligations, like the principle of the *restitutio in integrum,* are deemed to be part of the 'General Part of the Law of Obligations'. Rules that expound the peculiarities of remedies typical of certain specific obligations—such as the reduction of the purchase price as a remedy available to the buyer or the

[59] On those writers who adopted the system for no other reasons than that it was a well-entrenched tradition see Schwarz, above n 53, at 580. On the 'secular agnosticism' of the nineteenth century legal scientists and their refusal to make moral commitments see, eg, J Gordley, *The Philosophical Origins of Modern Contract Doctrine* (Oxford, Clarendon Press, 1991) 227.

[60] On 'will theory' thus understood as synecdoche for the nineteenth century 'legal consciousness' see D Kennedy, 'Two Globalizations of Law & Legal Thought, 1850–1968' (2003) 36 *Suffolk University Law Review* 631, 637. On how far the idea that this 'legal consciousness' is rooted in Kantian or Hegelian philosophy is still a matter of contention, see, for an introduction, H Dedek, *Negative Haftung aus Vertrag* (Tübingen, JCB Mohr (Paul Siebeck), 2007) 101. See also Gordley, *ibid*, at 227.

[61] Savigny, above n 58, at vol 1, § 52, 333: '[…] dass dem individuellen Willen ein Gebiet angewiesen wird, in welchem er unabhängig von jedem fremden Willen zu herrschen hat.'

[62] For an example of the outrage that was provoked when provisions on consumer protection were implanted in the General Part, see W Flume, 'Vom Beruf unserer Zeit für Gesetzgebung' [2000] *Zeitschrift für Wirtschaftsrecht* 1427.

payment of damages in periodic installments in case of a delictual injury that led to bodily harm—would be found in the context of the respective nominate contracts or the sections on delicts.

This scatteredness of the rules on damages draws our attention to another problem the undertaking of inducting general principles from particular rules faces. Abstract general principles do not apply in the abstract. General principles apply to particular cases. Damages are awarded if the breach of an obligation causes loss. However, the obligation in question is never just the pure, abstract and *universal* concept of an obligation stripped of all its worldly *accidentia*,[63] but is a *contractual* obligation or a *tortious* obligation or a *restitutional* obligation, and so on. If we are trying to distill general principles from all these different fields, we are running the risk that the common denominator is so insignificant, the general principle so abstract that it is actually too abstract to be applied in real life without adjustments to the respective context. The consequence is that, if a set of rules is classified as 'general', there is a need for additional rules that assure the adjustments of the general rules to the specific contexts of sales, torts, unjustified enrichment and so on.

Smith encounters the same difficulty when reflecting on a general principle of responsibility: surely, there is a general idea that underlies rules on the vitiation of contractual *consensus* as well as the rules on the responsibilities for one's wrongdoing. The concept of responsibility may be general, but 'duress and other general defences do not operate identically across contractual and non-contractual settings'.[64] There obviously is a theme of responsibility pertaining to both the law of contractual and non-contractual obligations, but its actual emanations in both fields are so different that it seems impossible to formulate a general rule—hence Smith's scepticism as to the classifiability of this principle of responsibility as belonging to the 'General Part of the Law of Obligations'.

The German Civil Code, keener on creating 'General Parts' wherever possible, has to deal with exactly this effect of creating repetitions instead of avoiding them.[65] For every general principle there have to be many other rules that adapt the general principle to special circumstances. The result is

[63] At least in regard to applying the rule to an actual factual situation; however, this is not necessarily meant to imply that, in the Nominalist sense, such a concept could not *exist* at all.

[64] Smith, above n 31 in the text accompanying fns 46 and 47.

[65] This malaise becomes particularly apparent when analysing the German Civil Code, but is, of course, a problem other codifications have to put up with as well. See, eg, the rules on 'lesion' in the Quebec Civil Code: Art 1406 ss 1 CCQ 1 defines 'lesion' as a principle, ss 2 adjusts the principle for cases involving minors or 'protected persons of full age.' Art 1405, however, explains that 'lesion' vitiates consent exclusively in respect to minors, to said protected persons, and the cases expressly provided by law. These cases are limited to extraordinary situations as, eg, addressed in Art 424. Against this backdrop, expounding the 'general principle' seems of rather limited value and can only be explained by its historical roots. See P-G Jobin, *Les Obligations*, 6th edn (Cowansville, Yvon Blais, 2005) 267ff.

that provisions dealing with the same subject matter are scattered all over the Code, as we just saw in the case of the rules on damages. This makes it very difficult to get a coherent picture of the law: the stones of the mosaic have to be pieced together somewhat tediously to be able to see a coherent picture. Or, if we want to return to the imagery of the clockwork: the German Civil Code turns out to be a veritable *'Grande complication'* whose mechanism only a most accomplished master clockmaker could understand in its full complexity.[66]

The recent reform of the German law of obligations[67] has even further increased this effect. In an attempt to erase anachronistic peculiarities like the 'Aedilitian' remedies available to the buyer in sales contracts (remnants of the Roman law of market transactions, administered by the magistrates in charge of market exchanges, the *aediles*, not the *praetor*[68]), remedies across all nominate contracts were supposed to be harmonised and, therefore, the importance of the 'General Part of the Law of Obligations' heightened. The technical means employed was a web of complex cross references cast over the Code, referring to the 'General Part of the Law of Obligations;'[69] in turn, another set of rules had to be introduced to adjust the general rules to the context of the different types of contracts.

Take, for example, a buyer's claim for damages 'in lieu of performance' in case of the sale of a defective good. Let us assume that a contract has been concluded according to the rules on offer and acceptance in § 116 BGB and the following paragraphs, that this contract is a contract of sales as defined by § 433 BGB, and that the good delivered is defective according to the definitions of § 434 BGB.[70] The buyer's claim for damages, her 'subjective right', would be seen as rooted in § 281 BGB, a 'general' rule on damages for breach of an obligation, which is entitled 'Damages in lieu of performance for non-performance or failure to render

[66] See also Zimmermann, above n 50, at 31.

[67] For a comprehensive introduction, see R Zimmermann, *The New German Law of Obligations* (Oxford, Oxford University Press, 2005).

[68] See Zimmermann, above n 50, at 311ff. For the reconstruction of the edict see O Lenel, *Das Edictum perpetuum: Ein Versuch zu seiner Wiederherstellung*, 3rd edn (Leipzig, Tauchnitz, 1927) 554ff; É Jakab, *Praedicere und cavere beim Marktkauf—Sachmängel im griechischen und römischen Recht* (München, CH Beck, 1997) 123ff, 153ff, 272, 291ff, 296.

[69] See, eg, B Dauner-Lieb, 'Die geplante Schuldrechtsmodernisierung' [2001] *Juristenzeitung* 8, 12ff, 16; H Dedek, 'Vorbemerkung zu §§ 241ff' in M Henssler and F Graf v Westphalen (eds), *Praxis der Schuldrechtsreform,* 2nd edn (Recklinghausen, ZAP-Verlag, 2002), *passim*.

[70] § 434 BGB partly incorporates the standards set by the Directive 1999/44/EC of the European Parliament and of the Council of 25 May 1999 on certain aspects of the sale of consumer goods and associated guarantees (OJ L 171 of 7 July 1999 at 12), generalising the definitions meant for the sale of consumer goods for all sales contracts.

performance as owed',[71] and whose purpose is to decide whether the creditor of the breached obligation has access to damages on the basis of an expectation measure, replacing her initial right to specific performance. § 437 Nr. 3 BGB, entitled 'Rights of buyer for defect,' forms the joint between 'general' and 'special' parts and declares that § 281 BGB is applicable as well in cases of sale of defective goods. § 437 Nr. 3 BGB explicitly refers to § 281 BGB, which is, in turn, a modification of the principle stated in § 280 BGB, which announces that damages have to be paid for breach of obligations, but only if the breaching party was at fault. To find out whether this is the case, we now need to look at the scope of obligations of the seller under a sales contract, which is again defined in the 'special part', in § 433 BGB. If we find that indeed an obligation has been breached, we jump back to § 276 BGB to consider the standards that decide whether the seller was at fault. If the requirements of § 280 BGB are thus satisfied, we return to § 281 BGB in order to check whether its next important requirement has been met: in order to replace the right to specific performance with a claim for damages, the creditor needs to give the debtor a warning and a second chance to perform properly within a reasonable timeframe. In certain cases, enumerated in § 281 section 2 BGB, such a warning and notice is dispensable, for example, in cases of fraudulent misrepresentation[72] or repudiation. However, if none of these exceptions apply, we have to—as if following a tennis match—shift our gaze back to the special rules on sales contracts again. There, we find § 440 BGB, which explicitly extends the list of exceptions expounded in § 281 section 2 BGB in order to adjust it to the specific interests of the parties to a sales contract. If we, finally, come to the conclusion that the buyer is indeed entitled to damages, the *quantum* would be calculated according to the guidelines found in the 'General Part of the Law of Obligations' again: § 281 BGB, ordering the expectation measure of damages 'in lieu of performance,' and § 249 BGB and following paragraphs, containing the rules around the general principle of *restitutio in integrum*.

Confusing as it might seem, the way I described the back and forth between the rules in the different parts of the Code is a gross oversimplification compared to the way a student—and eventually, a judge—would be expected to actually treat the matter.[73] Rather than sublime clockwork,

[71] Translations of the now official titles of the respective provisions as offered on the official website of the German Federal Ministry of Justice: <http://www.gesetze-im-internet.de/englisch_bgb/englisch_bgb.html> (last accessed November 2008).

[72] Bundesgerichtshof, *Beschluss vom 8. 12. 2006 – V ZR 249/ 05* (2007) Neue Juristische Wochenschrift 835 citing Dedek, above n 69, at § 281, [36].

[73] It would also be simple to present far more intricate and complex examples of this technique.

this mechanism seems more like a mental pinball machine that tosses you back and forth relentlessly between its special and its general parts.[74]

III. THE HARDENING OF CATEGORIES: THE CASE OF RELIANCE INTEREST IN CONTRACT DAMAGES

We have seen, so far, some problems that might arise in a system relying too heavily on a division of law into categories of 'special' and 'general' The 'atomisation' of law in such a system, however, is beyond just a technical problem that makes the application of rules more tedious. The ('vertical') borders between the different 'parts' do not just place certain rules in certain 'areas', they create intellectual boundaries as well that, at least over time, prevent jurists from seeing coherence and continuity above and beyond the borders. Seeing these particular connections can be as important as the ability to understand the generality of certain ideas. While categorising, it is therefore vital to keep in mind the permanent possibility of a different categorisation, and thus to keep open the recourse to different ways of legal reasoning. What we have to fear about categorisation in terms of 'general' or 'special' is not so much the possible incorrectness of such a categorisation, but its petrifaction—the 'hardening of categories' that cuts off the access to valuable legal arguments.

Stephen Smith concludes that 'failing to pay attention to contract law's vertical limits can lead scholars, judges, and lawyers to apply the wrong principles to understand legal rules, to draw the wrong inferences from those rules, and to fail to make appropriate generalisations'.[75] One could reply, from the perspective of a legal system built on rigorous classificatory schemes, that this might go both ways: taking these limits too seriously can lead scholars, judges and lawyers to fail to draw appropriate connections between the concepts separated by the vertical borders. The important and complicated task would be to keep both perspectives, limited and borderless, at the same time.

Let me try to illustrate my point by taking a closer look at the phenomenon known as the 'reliance interest in contract damages'. The article so entitled by Lon Fuller and William Perdue[76] is famed as being probably 'the most influential single article in the entire history of modern

[74] For a general criticism of the complexity of the German Civil Code, see Zweigert and Kötz, above n 40, at 144ff. On the relationship between the 'General Part' and the pervasive fascination for abstract concepts, the hallmark of German 'conceptual jurisprudence,' see, most recently, Stagl, above n 58, at 39.

[75] Smith, above n 31 in the first paragraph of Section VIII Conclusion.

[76] LL Fuller and WR Perdue Jr, 'The Reliance Interest in Contract Damages: 1' (1936) 46 *Yale Law Journal* 52.

contract scholarship, at any rate in the common law world'.[77] Interestingly, the Fuller and Perdue article is being hailed particularly (among others) for its *taxonomical* achievements. For example, Grant Gilmore, in language remarkably suitable to our exploratory theme, applauds Fuller for having 'discovered' a 'case law underground' that turned out to provide the protection of what he christened the 'reliance interest'.[78]

According to Smith's map, however, this famous article, by focusing on damage rules, treats a subject matter that lies outside the realm of contract law.[79] Even though a theory of the law of damages might, indirectly, provide helpful insights for the law of contracts as well,[80] these territories are separated by a 'vertical border'. This train of thought brings us, paradoxically, to another thrust of the article that was 'taxonomical' in nature as well. Fuller and Perdue attack, from the outset, the understanding that, when we evaluate contract damages by applying an expectation measure, we are just applying a *general rule* guiding the award of compensatory damages, namely, the rule that the victim of an injury simply is to be put in the same position, 'as far as money can do it', as if the injury had not occurred. The article calls this view into question; it tries to prove that the way we first formulate and then apply a 'general rule' of compensation to the situation of a breach of contract is already an expression of what we think the purpose of contract law should be. In now famous words, the authors wrote:

> Yet in this case we 'compensate' the plaintiff by giving him something he never had. This seems on the face of things a queer kind of 'compensation.' We can, to be sure, make the term 'compensation' seem appropriate by saying that the defendant's breach 'deprived' the plaintiff of the expectancy. But this is in essence only a metaphorical statement of the effect of the legal rule. In actuality the loss which the plaintiff suffers (deprivation of the expectancy) is not a datum of nature but the reflection of a normative order.[81]

This highlights the structural interlock between awarding compensatory damages according to the expectation measure and the purposes of contract law. We cannot just say there is institutionalised reliance in the

[77] PS Atiyah, *Essays on Contract* (Oxford, Clarendon Press, 1986) 73.

[78] G Gilmore, *The Death of Contract* (Columbus, Ohio State University Press, 1974) 56. See also TD Rakoff, 'Fuller and Perdue's *The Reliance Interest* as a Work of Legal Scholarship' [1991] *Wisconsin Law Review* 203, 206ff, who criticises Gilmore for stylising Fuller's achievement as 'discovery', and instead points to the intellectual groundwork done by Williston and Cohen.

[79] Smith, above n 31, text accompanying fns 11–14: 'Often regarded as the most famous article in English on contract law, Fuller and Purdue's 'The Reliance Interest in Contract Damages' focuses, as the title states, on damages rules. Contract courses in common law jurisdictions frequently begin with a discussion of damages. … But the rules governing contract damages are not contract law rules.'

[80] Smith, above n 31, in text accompanying fn 26.

[81] Fuller and Perdue, above n 76, at 53.

expectancy that justifies the treatment of an obligation as a property right, leading to the treatment of the loss of expectancy as an actual injury, when on the other hand this reliance is grounded in the very fact that enforcement of contractual promises *consists of awarding the expectancy*. In *taxonomical* terms, by showing this circular interdependence, 'The Reliance Interest in Contract Damages' tries to break down the conceptual boundaries between the 'general law of damages' and 'contract law'; the article seeks to *de-generalise* our understanding of the 'general' rules that purport to be applicable to actual tortious injuries as well as to the 'loss' suffered as a consequence of the breach of a contractual obligation. Its purpose is to make us think about the award of damages in contract cases as lying at the very heart of contract law and to dispense with any vertical borders that might separate the law of contract damages from the realm of contract law.

Now, this might be, in terms of the doctrinal implications, a good idea or not. This is not what is important for our inquiry. My point is that only in a system that is not founded on a deeply entrenched understanding of law as fractioned into 'general' and 'special parts', is it at all possible to see the kind of coherence and interdependence that Fuller and Perdue do. Only in such a system does the legal analyst have the freedom to transcend boundaries and recombine elements that one might see as separated by vertical borders; the freedom to go even one step further and proclaim that, as, for example, Patrick Atiyah did, the binding force of promise is grounded in reliance alone and that, therefore, the enforceability of promises should express itself in the award of reliance damages as the primary remedy.[82]

Again, the point is not whether this way of thinking has its doctrinal merits or not.[83] The point is that making such a connection would be literally *unthinkable* within the partitioned framework of German civil law.[84] Under the German approach of strict 'vertical' separation, 'contract' as a coherent institution is completely dissolved in the system. Formation of contract is a matter of the 'General Part'. Once the contract is formed, it gives birth to obligations, which are dealt with by another part of the mechanism, the 'General Part of the Law of Obligations'. Here, the obligation is seen abstracted from its contractual origin and treated as an ideal concept without an individual history. In case of a breach, the general rules on damages apply (as we have seen, possibly modified by more

[82] PS Atiyah, *Promises, Morals, and Law* (Oxford, Clarendon Press, 1981) 36ff; Atiyah, above n 77, at 20ff.

[83] For a discussion of this question, see SA Smith, *Contract Theory* (Oxford, Oxford University Press, 2004) 413ff; SA Smith, '"The Reliance Interest in Contract Damages" and the Morality of Contract Law', Issues in Legal Scholarship, Symposium: Fuller and Perdue (2001): Article 1, <http://www.bepress.com/ils/iss1/art1> (last accessed November 2008).

[84] On the following, see Dedek, above n 60, at 6ff, 111ff.

specific rules). The taxonomic scheme itself *prescribes* the application of the general rules on damages, thus equating loss of expectancy in contractual and delictual constellations. The question of what amount of damages is awarded becomes a mere mathematical operation based on the idea of causation, comparing the claimant's actual financial position after the injury to the hypothetical financial position he would be in but for the injury. The amount by which the latter exceeds the former is deemed to be the pecuniary loss 'caused' by the injury. Applying this formula, an investment made in reliance on a promise which turns out to be useless when the promise is broken is, strictly speaking, not 'caused' by the breach: even if the contract had been performed, the investment would have been made and would have diminished the investor's patrimony nonetheless. In terms of tangible, pecuniary loss, therefore, the hypothetical financial position the claimant would be in but for the breach equals the actual position after the breach took place. The mere fact that the investment has been rendered futile has been seen as an intangible loss. Therefore, by mechanically applying this formula the 'contractual' measure of damages simply does not entail reliance damages.

The very idea that reliance invested at the stage of the formation of contract could have any implications for the calculation of damages is inconceivable; the system simply has decided that both stages have nothing to do with each other. Once the system—the structure of the Code and the jurisprudence built on this structure—has petrified, it cannot escape its self-imposed rigidity. It has to react by inventing new devices that add to the complexity of the mechanism, for example, the reliance-based but extra-contractual institute of '*culpa in contrahendo*'.[85]

Courts, working with the formula that gains prevented by the breach are losses that can be compensated, have employed an institutionalised factual assumption that investments made in reliance would have been cost-efficient—explicitly pointing out that the futile investment as such is not the loss that has been incurred, but the loss of opportunity to make a profit that would at least have equalled the amount of the investment.[86] Because this differentiation was taken very seriously, new problems arose in cases where an actual intention to make a profit could not be shown.[87] It was held that in such cases the expectancy could not be measured by taking recourse to out-of-pocket investments and, therefore, no damages could be claimed at all.

Finally, the legislator reacted by implanting a new provision into the Code, § 284 BGB, granting a right to reliance damages. But even when the

[85] See now §§ 311 ss 2, 280 BGB.

[86] Bundesgerichtshof BGHZ 71, 234, 238ff; BGHZ 123, 96, 99: the so-called 'Rentabilitätsvermutung'.

[87] Bundesgerichtshof BGHZ 99, 182, 196ff: the so-called 'City Hall Case'.

legislator himself sprang into action, German legal scholars claimed that a provision granting reliance damages for breach of contract would be a monstrosity because it would violate the 'general law of damages' (as if it were a law of nature), particularly the general rule that the aggrieved party could only be compensated for losses that were caused by the breach, whereas this causal link was missing between the breach and the detrimental change in position incurred in reliance on the promise.[88]

Obviously, the problems related to causation have occurred to common lawyers as well: 'The reliance interest fixes the wrong as *making* the promise, not *breaking* it.'[89] It is understood that by claiming reliance damages, the aggrieved party is asking 'not to be put into the position he would have been in had the contract been performed, but to be put into the position he would have been in had it never been made.'[90] How is it, then, possible to award reliance damages for breach at all? It has been suggested that the award of reliance damages could be explained by the assumption that the disappointed party would have at least 'broken even'[91]—which does not, however, account for cases like *McRae v Commonwealth Disposals Commission*, where, in addition, recovery for the lost opportunity to profit from an alternative contract was allowed.[92] Without endorsing the extreme views put forward by Atiyah, the common law has still found its way around this seemingly 'logical' hurdle simply by being less 'dogmatic' than the German civil law. Judges help out by skating over this rough patch by stating that one who fails to perform his contract is justly bound to make good all damages that accrue or flow 'naturally from the breach'.[93] In *Anglia Television Ltd v Reed*, Lord Denning simply asked whether the loss—investments rendered futile by the breach—had been within the contemplation of the parties and granted, without much ado, redress for the loss of investments incurred before the contract was even entered into.[94] Such rulings could be seen as the result of the judiciary being too little concerned with proper doctrinal analysis. However, these

[88] H Altmeppen, 'Untaugliche Regeln zum Vertrauensschaden und Erfüllungsinteresse im Schuldrechtsmodernisierungsentwurf' [2001] *Der Betrieb* 1399, 1403ff; H Altmeppen, 'Nochmals: Schadensersatz wegen Pflichtverletzung, anfängliche Unmöglichkeit und Aufwendungsersatz im Entwurf des Schuldrechtsmodernisierungsgesetzes' [2001] *Der Betrieb* 1821, 1823.

[89] MB Kelly, 'The Phantom Reliance Interest in Contract Damages' [1992] *Wisconsin Law Review* 1755, 1775, n 60.

[90] H McGregor, *McGregor on Damages,* 17th edn (London, Sweet & Maxwell, 2003) [2–020].

[91] See, eg, EA Farnsworth, *Contracts*, 3rd edn (New York, Aspen Publishers, 1999) § 12.16, 837 ('Effect of losing contract'); JD McCamus, *The Law of Contracts* (Toronto, Irwin Law, 2005) 835.

[92] [1951] 84 CLR 377 (HCA).

[93] *Security Stove & Mfg Co v American Rys Express Co* 227 Mo App 175, 183 (1932), citing *Hobbs v Davis* 30 Ga 423, 425 (1869).

[94] *Anglia Television Ltd v Reed* [1972] 1 QB 60 (CA).

rulings also show how the system leaves its judges room for the reasonable exercise of discretion—a system that has never truly believed in technical perfection as the answer to the quest for justice.

Ironically—and counter-intuitively—it is the French civil law that has made the least efforts to lock in its judges by developing complicated theories around reliance and expectancy measure. It gets along without both categories altogether, relying simply on the medieval distinction between *damnum emergens* and *lucrum cessans*, the loss sustained and the gains prevented (see Article 1149 Code Civil (CC); Article 1611 Civil Code of Quebec (CCQ)). Since the times of Cujas, out-of-pocket losses are simply seen as *damnum emergens*:[95] 'Le calcul de la perte subie est relativement facile à faire: ce sont les dépenses effectuées et devenues inutiles.'[96] The seemingly obvious lack of a causal nexus between breach and loss is not a matter of concern.[97] The absence of the conceptual categories (expectation/reliance damages) has been described in a way that makes it seem as if French law has simply not yet reached the level of conceptual refinement to fully appreciate the distinction between expectation and reliance interest.[98] On the other hand, French civil law has not faced the problems that arose from the conceptual petrifaction the German system imposed on itself. Quite undogmatically, the French civilians simply embraced the intuition that a breach 'turns' an investment made in reliance into detriment, and therefore 'causes' a loss.

The idea that compensation for this 'actual' loss is an even more suitable and 'natural' reaction to the breach of promise was unfolded by Fuller and Perdue. By drawing on Aristotle, they demonstrated that compensation for a detrimental change of position in reliance on a promise is a matter of corrective justice and therefore represents a more pressing case for relief than the disappointment of having been deprived of the expectancy.[99] German civil law, however, locked itself in by mechanically applying the

[95] Iacobus Cuiacius, *Opera Omnia In Decem Tomos Distributa, Opera Postumae Quae De Iure Reliquit, Tomus Quartus*, Pars Prior, Paris, 1658, Comment. in Tit. I. de act.emp et vend. Lib. XIX Dig., ad l. 21, §.Cum per venditorem, Col 822: *Dicimus autem circa rem ipsam, consistere pretium rei quod emptori abest, quod praenumeravit venditori, & etiam quod emptionis causa erogavit, puta, quod dedit parario,* au courretier: *quod dedit uxori (ut fit nonnumquam) venditoris, quod dedit praeconi aut quaestori, quo interveniente facta auctio est, quod dedit fisco vectigalis nomine.* [We, however, say that the interest relating to the object of sale itself [circa rem ipsam] comprises the value of the thing the buyer has been deprived of, the deposit the buyer gave to the seller, and also the expenses he made because of the contract of sale, eg, the amount he paid to the realtor, or the wife of the seller (as it sometimes happens), or to an auctioneer or a magistrate to avoid a sale by autction, the taxes he paid to the authorities.]

[96] Jobin, above n 65, at 905.

[97] G Ripert, *Traité de droit civil*: *d'après le traité de Planiol* (Paris, Librairie générale de droit de jurisprudence, 1956) vol 2, 247.

[98] GH Treitel, *Remedies for Breach of Contract* (Oxford, Clarendon Press, 1988) 89.

[99] Fuller and Perdue, above n 76, at 56.

'general rules on damages'. Ideas that do not fit the grid because they transcend the 'vertical' borders have no place in legal discourse: thus the 'hardening of categories' constricts the arsenal of legal reasoning.

IV. CONCLUSION

Observed from a comparative perspective, the attempts at 'mapping' the common law trigger ambivalent reactions. Particularly for civilians, nothing seems more alluring than a more coherent map of the common law[100] which would help the outsider to overcome his confusion when trying to find his way around. On the other hand, one is tempted to extend a warning: Be careful what you wish for! The civil law, and especially the German expression of the civilian tradition, has experienced the downside of strict systematisation and classification. This temptation to warn might also have to do with the grass always seeming a bit greener on the other side—while both civilians and common lawyers (more or less openly) tend to think of their respective 'systems' as somewhat superior (more orderly, more efficient, and so on), the harmony and order of the civil law does not fail to appeal to the common lawyer, whereas the freedom and flexibility of common law reasoning tempts the civilian, in particular the academic.

I have emphasised, however, that it was not the purpose of this article to criticise the intellectual merits of the 'mapping' enterprise. I simply tried—for purposes of comparison and, by all means, deterrence—to highlight some problems that arise if a taxonomic approach to law is pushed to its extreme, as it has been in the German civil law tradition. One problem arising in this context is that of excessive mechanical complication. The German map of private law is so overly sophisticated that it, in all its complexity, really opens up only to the most ardent adept of German 'Legal Science.' In a world where legal systems and styles are said to be competing, this characteristic will—particularly in the light of the complete obliviousness of American law to such a kind of jurisprudence—be more and more of a liability.

The foremost risk of taxonomy and classification in law, however, has turned out to be the 'hardening of categories'. Simply because taxonomy in law is not just descriptive, but prescriptive and normative in nature, one must be always willing, once a classificatory scheme has been conceived, to call it into question again. Stephen Smith acknowledges this necessity by integrating into his twofold test the question of whether a classification

[100] This yearning to understand and to appropriate also underlies the attempts of comparative functionalism to draw a global map and to develop trans-systemic categories of contractual concepts, as, eg, the famous category of the 'Indicia of Seriousness,' see Zweigert and Kötz, above n 40, at 388ff.

makes moral sense. This part of the test safeguards against the allure of a 'logical' or 'scientific' classification that purports to be above and aloof from current ideas and moralities.[101] Becoming aware of this 'moral' implication of legal taxonomy leads to the insight that we must not accept traditional doctrinal categories as givens simply for their 'logical' appeal. On the other hand, if taxonomy has to account for its making 'moral' sense, naturally the question arises, what *does* make 'moral' sense? Having to answer this question as well makes the ambition of bringing a bit more 'logic' to the common law, a truly Herculean task.[102]

Finally, the 'hardening of categories' once engrained might not just lead to the encapsulation of possibly outdated ideologies, but also, as I have tried to show, to an intellectual rigidity that curtails the argumentative potential of legal discourse. The discourse censors ideas that do not fit the grid of *quasi-natural* divisions and demarcation lines. However, every cartographer of knowledge must recall that '*the map is not the territory*'.[103] My conclusion would therefore be this. It could be reasonable and intellectually stimulating to redraw the map of contract law by establishing 'vertical borders'; however, keep such frontiers open so that ideas can still migrate freely. Beware of the petrification of categories. Classify with care; always taxonomise in moderation.

[101] Which is, of course, even more problematic if such ideas have crystallised in the structure of a Civil Code, which is, at least in France and Germany, seen as a national treasure, and therefore tends to be fiercely guarded from change. On the attempts to initiate a reform of the French Law of Obligations see, for an introduction, E Hondius, 'The Two Faces of the Catala Project—Towards a New General Part of the French Law of Obligations' (2007) 15 European Review of Private Law 835–9.

[102] See, eg, P Birks, 'Equity in the Modern Law: An Exercise in Taxonomy' (1996) 26 *University of Western Australia Law Review* 1, 5.

[103] A Korzybski, *Science and Sanity: An Introduction to Non-Aristotelian Systems and General Semantics,* 3rd edn (Lakeville, Connecticut, The International Non-Aristotelian Library, 1948) 58 (emphasis added).

3

Principle in Contract Law: the Doctrine of Consideration

STEPHEN WADDAMS

CHARLES ADDISON WROTE, in 1847, that

> The law of contracts may justly indeed be said to be a universal law adapted to all times and races, and all places and circumstances, being founded upon those great and fundamental principles of right and wrong deduced from natural reason which are immutable and eternal.[1]

Many other nineteenth-century writers spoke also of the 'principles' of contract law, including Frederick Pollock (1876)[2] and William Anson (1879),[3] the titles of both of whose books commenced with the word 'Principles.' These writers implied, both by the titles to their books and by remarks addressed to their readers, that there were certain propositions about English contract law, deserving of the name 'principles', that had some sort of special status as primary, fundamental, or indisputable, or that constituted a source from which rules used to determine particular cases were derived, and that those propositions could be identified and formulated.

But if we ask precisely what *were* these principles, the answer proves surprisingly elusive. Part of the reason for this is the indeterminacy of the word 'principle': the word has been used in many different senses, the meaning varying according to what is implicitly contrasted with it (principle and rule, principle and policy, principle and precedent, principle and authority, principle and pragmatism, principle and practice, principle and

[1] CG Addison, *A treatise on the law of contracts and rights and liabilities ex contractu* (London, W Benning, 1847) iv–v.

[2] F Pollock, *Principles of Contract at Law and in Equity* (London, Stevens and Sons, 1876).

[3] W Anson, *Principles of the English Law of Contract* (Oxford, Clarendon Press, 1879).

convenience, principle and utility, general principle and particular rule, general principle and particular case); on a controversial legal question two or more conflicting principles can usually—perhaps always—be identified; principles may be stated and restated at an infinite number of levels of generality, and commonly the word has been used to mean no more than a reason or rule framed at a higher level of generality than another. Sometimes a principle has meant more than a rule—a rule that is absolutely stringent, but at other times the word has signified something less than a rule—an objective desirable in general terms but liable to be outweighed by countervailing considerations. Whenever it is said—as it often is—that two principles come into conflict, one or both of them are liable to be outweighed by countervailing considerations. Often the meaning of the word merges with the idea of 'maxim.' Very commonly also it has signified a reasoned, or a well-reasoned legal argument; often it has meant a legal rule, or a reason in support of a rule, that the writer or speaker considers persuasive, legitimate, or satisfactory.

As we have seen, Addison asserted in 1847 that the principles of contract law were immutable, eternal and universal. This kind of assertion rests on a 'syncretic and ahistorical supposition'[4] that all law everywhere must be governed by the same principles. An examination of the doctrine of consideration, before and after 1847, shows that the conceptual bases of the doctrine, often, but not always, called principles, have varied markedly from time to time, as has the substance of the law. A historical study cannot establish what is the correct or preferable meaning of the word 'principle' nor what are the correct or preferable rules of contract law. But it can show that neither the usage nor the law has been immutable or eternal.[5]

About 80 years before Addison's statement William Blackstone had published his *Commentaries on the Laws of England*. He likened his work to a map, writing that 'an academical expounder of the laws . . . should consider his course as a general map of the law, marking out the shape of the country, its connexions and boundaries, its greater divisions and principal cities.'[6] Blackstone's map gave no definite place to contract law, though he mentioned contracts in several places. A definition of contract was offered in Book 2 (rights of things) as part of a chapter (entitled 'Of

[4] The phrase is David Ibbetson's: Sir W Jones, *An Essay on the Law of Bailments*, edited with introductions by David Ibbetson (Bangor, Welsh Legal History Society, 2004) 66. Syncretism, in theology, is the claim that all religions are fundamentally similar.

[5] As for the 'syncretic' claim that the principles of contract law have been universal, this would have to be tested by a comparison of English law with other legal systems. It seems implausible that the claim could be sustained in respect of the doctrine of consideration, which is absent from many legal systems.

[6] W Blackstone, *Commentaries on the Laws of England* (Oxford, Clarendon Press, 1765) vol 1, 35.

title by gift, grant, and contract') dealing with methods of acquiring rights to property.[7] In Book 1 (rights of persons) contracts were mentioned as part of the law of master and servant, and of husband and wife, and in Book 3 (private wrongs) assumpsit was mentioned as providing a remedy for breaches of promises, considered as wrongs. In respect of husband and wife, Blackstone wrote that 'our law considers marriage in no other light than as a civil contract'.[8] This assertion is startling at first sight—so much so that, when quoted, it has usually been to jocular or facetious effect, for the differences, in Blackstone's time, between marriage and other contracts were so many and so obvious that they scarcely require enumeration. But the question, for present purposes, is not what the assertion reveals about Blackstone's view of marriage, but what it reveals about his view of contracts. The comment shows that Blackstone did not visualise contracts as a body of controlling principles from which legal obligations were derived: if he had thought in those terms he would have given contract law a place on his map. It was not that Blackstone thought contracts unimportant—he refers to marriage as 'the most important contract of any'[9]—and contracts were an important means of transferring property rights, and an important aspect of the law of private wrongs. But he did not think of legal issues as 'part of' an independently existing contract law. It would be more accurate to say that he thought of contracts as 'part of' several different areas of the law—a means of effecting various kinds of legal consequences—and hence to be found in several different places on his map. The conception of contracts as a distinct body of principles was not, at this time, established in English law.

Often the idea of 'principle' has been used to denote the reason that is supposed to underlie a legal rule. One difficulty with this approach, from a historical perspective, is that it tends to dissolve the rule into the supposed principle. To take a simple example, the minimum age of contractual capacity at common law was 21 years. It may be supposed that the underlying reason for this rule was to ensure that contracting parties had sufficient understanding and maturity. But a legal rule exists independently of the reason for it, even when a single reason can be identified, and, of course, a rule that contracting parties must have attained the age of 21 years is not the same thing as a rule that contracting parties must be of sufficient understanding and maturity: there are many mature 20-year-olds, and many immature 21-year-olds. A more serious difficulty is that very often legal rules have been supported by several different reasons, which may be inconsistent with each other, and each one is insufficient, standing alone, to explain or support the rule. For example, two or more reasons

[7] *Ibid*, at vol 2, ch 30.
[8] *Ibid*, at vol 1, 421.
[9] *Ibid*, at 424.

might be given in support of a legal rule, none of which is sufficient in itself either as explanation or as justification, and which, though often coinciding in their legal result, tend to require opposite results in particular cases.

Nowhere is this phenomenon more apparent than in relation to the rule that a contractual promise must be supported by consideration. In an unpublished treatise written about 60 years earlier than Blackstone's *Commentaries*, Sir Jeffrey Gilbert, later chief Baron of the Exchequer, like Blackstone in his second book, envisaged contract primarily as a means of transferring property. Gilbert commenced his treatise by saying:

> Contracts are two-fold: verbal and solemn. Now contract is the act of two or more persons concurring, the one in parting with, and the other in receiving some property right or benefitt. The most notorious way of transferring of right from one to the other is this by contract for all men by their labour and industry did first acquire to themselves a property so they may by other acts of their own transferr that property where they please, and all laws have allowed it as a settled maxim that the right of disposall must of necessity follow the rights of absolute dominium, for certainly as a man may be industrious for himself he may be so also for another and therefore the establishment of the propriety [*sic*] must be in his hands to whom the disposition is made and no doubt as the notion of propriety was begotten from humane necessity so was also this of contract.[10]

Gilbert then devoted many pages to the topic of consideration, with minute discussion of numerous hypothetical and decided cases, showing that consideration was regarded by him as an important, difficult, controversial and complex topic. As his first reason for the legal requirement of consideration, Gilbert gives the need to protect potential defendants from liability for rash promises. Having said that some opinions favoured 'the punctuall performance of every verbal promise', he continued:

> Others held that no obligation arises from a naked promise and that the force of the engagement doth totally depend on the consideration and they take it to be a thing of great rigour that a man should dispose of the fruits and effects of a long and painfull industry and all the certain advantages and conveniences of life by the meer breath of a word and the turn of an unwary expression; they also think that the very laws of self-preservation will not permitt it for what reason of conscience can oblige a man to those words that tend to his own destruction, but if a valuable consideration had been received the bargain is compleat for another man's industry comes in the place of his own . . ..

Gilbert continued by saying that English law 'hath held the middle between these two extreames', in that formal contracts were enforceable

[10] J Gilbert, *Of Contracts* (*c* 1710) British Library, Hargrave 265, f 39 (some punctuation added, abbreviations expanded, and capitalisation removed).

> so that if a man will oblige himself under the solemnitys of law whereby his contract appears to be seriously intended, it shall ever be obligatory and the consideration shall be intended . . . but if the contract be verball only it binds in respect of the consideration, otherwise a man might be drawn into an obligation without any real intention by random words, ludicrous expressions, and from hence there would be a manifest inlet to perjury because nothing were more easy than to turn the kindness of expressions into the obligation of a real promise.[11]

'Consideration' is a conveniently flexible word, embracing three very different ideas: deliberation, reason for making the contract, and reason for enforcing it. Gilbert's reference to perjury adds a fourth idea: the requirement of consideration tends to supply reliable evidence that the promise in question has actually been made.

This last idea was taken up by Lord Mansfield in *Pillans and Rose v Van Mierop and Hopkins*.[12] In that case a promise had been given in a commercial context by the defendant, in a signed writing, to guarantee repayment of money already advanced by the plaintiff. Lord Mansfield favoured enforcement. He said:

> I take it, that the ancient notion about the want of consideration was for the sake of evidence only; for when it is reduced into writing, as in covenants, specialties, bonds &c there was no objection to the want of consideration. And the Statute of Frauds proceeded upon the same principle. In commercial cases amongst merchants, the want of consideration is not an objection.[13]

Wilmot J thought that the requirement of consideration originally 'was intended as a guard against rash inconsiderate declarations'.[14] This reason also would have been sufficient to justify enforcement in the particular case, but he and the other judges managed, with some difficulty, to find consideration (in the sense of value exchanged). Subsequent references to the case have tended to amalgamate the principle favoured by Lord Mansfield (promises are enforceable if there is reliable evidence that they were made) with that favoured by Wilmot J (promises are enforceable if made with due deliberation). The proposition that promises are enforceable if made with serious intent, for example, tends to fuse the two ideas, and a requirement of writing might satisfy both simultaneously,[15] but the ideas are distinct as Lord Mansfield's reference to the Statute of Frauds reveals, for clearly there may be reliable evidence that a promise was in fact made, even though made rashly. Lord Mansfield, by appealing to a proposed principle, attempted to dissolve the rule that consideration was

[11] *Ibid*, at 39–40ff (some punctuation added, abbreviations expanded, and capitalisation removed).
[12] (1765) 3 Burr 1663, 97 ER 1035.
[13] *Ibid*, at 1038.
[14] *Ibid*.
[15] See L Fuller, 'Consideration and Form' (1941) 41 *Columbia Law Review* 799.

always required into some such rule as that promises in writing, or commercial promises in writing, were enforceable. Arguments can, of course, be made in favour of Mansfield's opinion, but it was not a historically accurate account of the previous law, and his view was decisively rejected by the House of Lords, on the advice of the judges, a few years later in *Rann v Hughes*.[16] Standing alone, the need to protect the promisor from rash promises could not explain the English law of consideration. Neither, standing alone, could the need for reliable evidence. But, as Gilbert's writing shows, both these ideas have been influential.

Included in Gilbert's account of consideration, quoted above, is another reason for the doctrine of consideration, namely, that it tends to prevent the dissipation of wealth by assuring to the promisor an equivalent in exchange for the wealth he gives up ('what reason of conscience can oblige a man to those words that tend to his own destruction, but if a valuable consideration has been received the bargain is compleat for another man's industry comes in the place of his own'). Standing alone, this reason also is insufficient to explain or to justify the actual law, because there was no need in English law for consideration to be of equal value to the promise sought to be enforced. But this is not to deny that the reason was, as a matter of history, influential, or that, as a matter of fact, it did (in some cases, though not in all) tend to prevent the dissipation of wealth.

Another set of interrelated reasons for the doctrine of consideration relates to the ideas of reciprocity and entitlement. As Guenter Treitel wrote, the claims of a promisee who has given nothing for the promise 'are less compelling than those of a person who has given (or promised) some return for the promise'.[17] One who has paid for a promise has a stronger claim to assert an entitlement to performance of the promise than one who has not paid for it, and the promisor has a correspondingly greater obligation to perform the promise, and to perform it to its full extent. One of the reasons that tend to support specific performance and the expectation measure of damages is that the promisee has bought and paid for the right to performance. Peter Benson has demonstrated the links among the concept of entitlement, the doctrine of consideration, and the extent of the usual remedies for breach of contract.[18]

Yet another reason that has been given in support of the doctrine of consideration is that it tends to protect creditors in cases where the promisor is insolvent. Lord Denman said, in *Eastwood v Kenyon*:

[16] 4 Bro PC 27, 2 ER 187. Summarised in (1778) 7 TR 350.

[17] GH Treitel, *The Law of Contract*, 11th edn (London, Sweet & Maxwell, 2003) 67.

[18] P Benson, 'The Unity of Contract Law' in P Benson (ed), *The Theory of Contract Law* (Cambridge, Cambridge University Press, 2001) 118.

> The enforcement of such promises by law, however plausibly reconciled by the desire to effect all conscientious engagements, might be attended with mischievous consequences to society; one of which would be the frequent preference of voluntary undertakings to claims for just debts. Suits would thereby be multiplied, and voluntary undertakings would also be multiplied, to the prejudice of real creditors. The temptations of executors would be much increased by the prevalence of such a doctrine and the faithful discharge of their duty be rendered more difficult.[19]

Where, as is common in cases of charitable pledges, the action is brought against the estate of the promisor after death, there is the added consideration that enforcement of the promise will be at the expense of possibly needy dependants.[20]

One reason commonly given for the enforcement of promises is that it tends to protect the reliance and expectation of the promisee. If, as just suggested, one who has paid for a promise has a stronger claim to enforce performance than one who has not paid, enforcement and consideration are necessarily interlinked. William Paley, in his *Principles of Moral and Political Philosophy* (1785) gave this as the primary reason for enforcing promises:

> Those who argue from innate moral principles suppose a sense of the obligation of promises to be one of them; but, without assuming this, or anything else, without proof, the obligation to perform promises may be deduced from the necessity of such a conduct to the well-being, or the existence indeed, of human society. Men act from their expectations. These expectations are in most cases determined by the assurances and engagements we receive from others. If no dependence could be placed upon these assurances, it would be impossible to know what judgment to form of many future events, or how to regulate our conduct with respect to them. Confidence therefore in promises, is essential to the intercourse of human life . . . But there could be no confidence in promises, if men were not obliged to perform them: the obligation therefore to perform promises is essential, to the same end, and in the same degree.[21]

Paley was not a lawyer, but he had a strong influence on nineteenth-century English thought, including legal thought. His *Principles* was for many years included in the very select reading list of the Cambridge Board of Moral Sciences Studies.[22] A long extract from his chapter on promises, with an elaborate example designed to show that promises were to be

[19] *Eastwood v Kenyon* (1840) 11 Ad & E 438, 113 ER 482, 487.

[20] In the leading Canadian case of *Governors of Dalhousie College v Boutilier Estate* [1934] SCR 642, 3 DLR 593 the promisor had suffered 'severe financial reverses which prevented him from honouring his pledge'. The action was brought against the estate after the death of the promisor.

[21] W Paley, *Principles of Moral and Political Philosophy* (London, 1785) 106.

[22] W Whewell, *Lectures on the History of Moral Philosophy in England*, new edn (Cambridge, Deighton Bell, 1862) 279.

interpreted in the sense understood by the promisee, if known to the promisor, and not in the sense subjectively intended by the promisor, was included in Joseph Chitty's much-used treatise on the law of contracts (2nd edition, 1834) and maintained in successive editions until nearly the end of the century.[23] Paley was expressly relied on in argument in the important case of *Smith v Hughes*,[24] and plainly influenced the court's formulation in that case of the objective principle of contractual obligations.[25]

Paley continued his discussion of promises by considering and rejecting the argument that society could manage satisfactorily without enforceability of promises:

> Some may imagine, that, if this obligation were suspended, a general caution, and mutual distrust would ensue, which might do as well: but this is imagined, without considering, how every hour of our lives we trust to and, depend upon others; and how impossible it is, to stir a step, or, what is worse, to sit still a moment without such trust and dependance. I am now writing at my ease, not doubting (or rather never distrusting, and therefore never thinking about it) but that the butcher will send in the joint of meat, which I ordered; that his servant will bring it; that my cook will dress it; that my footman will serve it up; and that I shall find it upon [my] table at one o'clock. Yet have I nothing for all this, but the promise of the butcher, and the implied promise of his servant and mine. And the same holds, of the most important, as well as the most familiar occurrences of social life.[26]

It is notable that all the examples given are of exchange transactions. If one thinks, as Paley did, of what promises are useful or necessary to an organised society, one thinks naturally of exchange transactions, not of gift promises. It is easy to imagine a well-organised society in which gift promises are not enforceable—eighteenth-century English society was, of course, a case in point—but it is more difficult to imagine an organised society without enforcement of exchange transactions. In this way, consideration has been linked with ideas of expectation and reliance, and, more generally, of social order.

A rule supported by so many and so various reasons leads inevitably to instances where the rule, as currently formulated, applies, but where the

[23] J Chitty, *A Practical Treatise on the Law of Contracts not under Seal: and upon the usual Defences to Actions thereon*, 2nd edn (London, S Sweet, 1834) 62, 12th edn (London, Sweet & Maxwell, 1890) 127–8. Not found in 1st edn (1826) or in 13th edn (1896).

[24] (1871) LR 6 QB 597.

[25] *Ibid*. Blackburn J said, at 607, '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.' Paley was expressly mentioned by Hannen J, at 610, for the important corollary that a promisor cannot be held to a meaning known by the promisee *not* to have been intended.

[26] Paley, above n 21, at 106–7.

underlying reasons, or some of them, do not. English law has had much difficulty with the question of modification of contracts, where additional value is promised by one party in exchange for performance by the other of an obligation already due under an earlier contract. It has often occurred that during performance of a contract circumstances change so that a party is in a position to demand from the other a higher payment than that originally agreed. A number of nineteenth-century cases involved sailors, who, having agreed to serve for the whole of a voyage at a certain wage, subsequently demanded a higher wage when the ship was at a place where substitute services were not readily obtainable. The renegotiated contracts were generally set aside. According to Campbell's report of the leading case of *Stilk v Myrick*, the reason for this result was that performance of a pre-existing contractual duty could not constitute consideration.[27] Thus, where a contract had been entered into at a fixed price, a subsequent renegotiation or variation consisting of an agreed increase in the price was unenforceable.

The rule in *Stilk v Myrick* (as reported by Campbell, and as it was generally understood to be) was much criticised on the grounds that it did not correspond to commercial understanding, that it failed to recognise that actual performance was of greater real practical value than a legal right to performance, and that it was easily circumvented by the parties or by a court desirous of enforcing the variation.[28] But criticism was often tempered with the observation that the rule, though difficult to defend in terms of consideration, was yet serving a useful purpose in offering, albeit indirectly, some legal protection against taking undue advantage of economic pressure. *Stilk v Myrick* itself was a case in point, where sailors, having agreed to serve on a voyage for certain wages, were promised higher wages in order to induce them not to desert during the course of the voyage. It is evident that one of the reasons for the decision was to protect the shipowner from a potentially extortionate threat by the crew to desert the ship in a distant place where there was no ready supply of substitute labour. In an earlier case, *Harris v Watson*, the court had also refused to enforce such a contract, but had given as the reason that

> if this action was to be supported, it would materially affect the navigation of this kingdom . . . for if sailors were in all events to have their wages, and in times of danger entitled to insist on an extra charge on such a promise as this, they would in many cases suffer a ship to sink, unless the captain would pay any extravagant demand they might think proper to make.[29]

[27] (1809) 2 Camp 317, 170 ER 1168.

[28] B Reiter, 'Courts, Consideration, and Common Sense' (1977) 27 *University of Toronto Law Journal* 439.

[29] (1791) Peake 102, 170 ER 94.

A different report (Espinasse) of *Stilk v Myrick*, running together the ideas of principle and policy, states that the judge (Lord Ellenborough) 'recognised the principle of the case of *Harris v Watson* as founded on just and proper policy',[30] and in a subsequent decision in the Admiralty Court (*The Araminta*, 1854) where the sailors had secured payment in gold of the extra money in advance of the extra performance (they were tempted to desert to gold diggings in Australia in 1852) and so the doctrine of consideration was of no assistance, *Stilk v Myrick* was interpreted as holding that the variation of the contract was 'illegal'.[31]

In *Williams v Roffey Bros & Nicholls (Contractors) Ltd* (1990) the English Court of Appeal held a renegotiation to be enforceable.[32] In that case, a subcontractor had contracted to perform carpentry work at an agreed price. When the work was partly done it became clear that the subcontractor would not complete it at the contract price, and the head contractor, who was subject to a penalty clause in the main contract for delay in completion, agreed to pay a higher price for the carpentry work. This latter agreement was held to be enforceable. The court held that performance of an existing obligation might constitute consideration. References to 'principle' were prominent. Glidewell LJ, who gave the leading judgment, rejected the argument that this conclusion was contrary to principle:

> If it be objected that the propositions above contravene the principle in *Stilk v Myrick* I answer that in my view they do not; they refine, and limit the application of that principle, but they leave the principle unscathed . . . it is not in my view surprising that a principle enunciated in relation to the rigours of seafaring life during the Napoleonic wars should be subjected during the succeeding 180 years to a process of refinement and limitation in its application in the present day.[33]

But Glidewell LJ added the very significant proviso that the renegotiation would be liable to be set aside if there was economic duress, which he called 'another legal concept of relatively recent development',[34] thereby suggesting that the result in *Stilk v Myrick* might be supported, though not on the reasoning given in Campbell's report. Many have welcomed the demise of consideration in this context, but it is not easy to say precisely what has replaced it. What exactly, in the court's view, was the governing concept of enforceability and how did it apply in practice to contractual

[30] 6 Esp 129, 130, 170 ER 851. Espinasse did not have a very high reputation as a reporter, but on the other hand he was one of the counsel in the case and had the means of knowledge.

[31] *The Araminta*, (1854) 1 Sp Ecc & Ad 224, 164 ER 130 (Adm Ct).

[32] [1991] 1 QB 1 (CA) 10, 19.

[33] *Ibid*, at 16.

[34] *Ibid*, at 13.

renegotiations? These questions are not very easily answered.[35] One approach has been to attempt to distinguish between a 'threat' and a mere 'offer to renegotiate', on the ground that there is something wrongful about a threat but not about an offer.[36] In *Williams v Roffey Bros* it was suggested that there was no duress because the proposal for renegotiation emanated from the head contractor. Purchas LJ thought this a conclusive point,[37] but it is evident that the subcontractor had indicated, by conduct if not by words, that he was not likely to complete the work on time at the contract price. It cannot be crucial that the threat not to complete was implicit rather than explicit. Indeed, it may be in the very cases where there is no real choice that it is unnecessary to spell out the threat, or even to make what could readily be called a 'threat' at all. In *The Araminta*, the case of the payment to the ship's crew at the Australian gold diggings in 1852, it was the master who, after several desertions, took the initiative and called together the rest of the crew, offering them increased wages if they would work the ship short-handed. Dr. Lushington said, of the master's payment, that it was made voluntarily, adding:

> I have used the expression *voluntarily*, because I think the effect of the evidence is, that the crew exercised no compulsion towards him, though, perhaps, in another sense of the word, such payment was not voluntary, and the more apt expression may be, and the one nearest the truth, that he was compelled by circumstances to make that payment.[38]

This is indeed often an apt expression to describe such circumstances, and for this reason it is doubtful whether the conclusion can be resisted that in *Williams v Roffey Bros*, as in most cases of this sort, there was a threatened breach of the first contract. The decision that the modified contract was nevertheless enforceable lends support to the view that, where the pressure on the other party is not excessive,[39] many courts have accepted that it is legitimate to gain an advantage in this way.

This leaves it very difficult to state what principles govern modification of contracts. Glidewell LJ said that the principle of *Stilk v Myrick* was left 'unscathed',[40] but any formulation of that principle before 1990 would

[35] For a fuller discussion see S Waddams, 'Commentary on "The Renegotiation of Contracts"' (1998) 13 *Journal of Contract Law* 199; S Waddams, 'Unconscionable Contracts: Competing Perspectives' (1999) 62 *Saskatchewan Law Review* 1.

[36] See S Smith, 'Contracting under Pressure: a Theory of Duress' [1997] *CLJ* 343.

[37] *Williams v Roffey Bros*, above n 32, at 21.

[38] *The Araminta* above n 31, at 133 (emphasis in original).

[39] *Restatement (Second) of Contracts* (1979) § 89 provides that modifications are enforceable if 'fair and equitable' in view of unexpected circumstances. The Uniform Commercial Code, s 2–209 makes modifications enforceable, explicitly subject however, by comment 2, to a test of good faith.

[40] Purchas LJ said, at 21, that 'the rule in *Stilk v Myrick* remains valid as a matter of principle, namely, that a contract not under seal must be supported by consideration', but the

have plainly demanded the opposite result to that reached in *Williams v Roffey Bros*.[41] Glidewell LJ was consciously changing the law as it had previously been understood to be, but he found it necessary to say that the 'principle' remained intact. Though 'unscathed' (that is to say, undamaged) the principle is said to be refined and limited, but it is not at all clear what these refinements and limitations amount to. Consideration is still necessary; however, performance of a pre-existing contractual duty may amount to consideration (but does not always do so); however, again, even if there is consideration the modification will not be enforceable if there is economic duress; however, yet again, it is not explained why the quite severe pressure on the head contractor in *Williams v Roffey Bros*. (the threat of the penalty clause) did not amount to duress. Breach of contract is a wrong, but apparently it is permissible to gain an advantage by threatening to break a contract. Economic duress was said, by Glidewell LJ, to 'provide another answer in law to the question of policy which has troubled the courts since before *Stilk v Myrick* and no doubt led, at the date of that decision, to a rigid adherence to the doctrine of consideration'.[42] 'Rigid,' in legal argument, is never a word of approbation, so the suggestion seems to be that *Stilk v Myrick* can be justified only on the basis of 'policy', but it is not very clear what the policy was or is (protection of the navigation of the Kingdom, or avoidance of economic duress), or how it relates to the 'unscathed' principle.

Another aspect of the doctrine of consideration has been the troublesome question of contracts for the benefit of third parties. In the old case of *Dutton v Poole*, a father, wishing to give money to his daughter and proposing to cut down trees to raise the money, agreed with his son and heir that he would refrain from cutting down the trees if the son would pay the daughter £1,000.[43] The son inherited the land with the timber intact, but refused to honour his promise. The promise was held to be enforceable.

Until the mid-nineteenth century it was generally accepted by writers that this case was good law, though it was evidently an exception to the

proposition that a contract must be supported by consideration is not the 'principle' or the 'rule' for which *Stilk v Myrick* had been cited during the previous 180 years; Purchas LJ seems to be suggesting, in view of the conclusion in *Williams v Roffey Bros*, that, though the 'rule' in *Stilk v Myrick* 'remains valid as a matter of principle', the principle must have been misapplied in *Stilk v Myrick* itself.

[41] For example, GC Cheshire and CHS Fifoot, *The Law of Contract*, 6th edn (London, Butterworths, 1964) 77, calling it 'the somewhat obvious rule, that there is no consideration if all that the plaintiff does is to perform or to promise the performance of an obligation already imposed upon him by a previous contract between him and the defendant'.

[42] Above n 32, at 14.

[43] (1689) 2 Lev 210, 83 ER 523, affirmed T Raym 302, 83 ER 156 (Ex Ch).

idea of privity of contract.[44] A variety of explanations was proposed. Gilbert approved of the result of the case, but encountered some difficulty with formulating a principle and defining its limits:

> So in consideration that the father of the defendant whose heir he is would not cut trees to pay the portion of the wife of the plaintiff sister of the defendant he assumed to pay to the wife of the plaintiff 1000*l*; the daughter and her husband may have an action on her father by reason of the nearness of relation and it is a debt to the daughter to make a provision [portion?] for her and so she is interested in the consideration and the son hath benefit by it; otherwise it is if the money had been paid to a stranger; the law therefore will put a near relation in the place of the promisee because the promisee is bound by the law of nature to provide for such relation and therefore the labour of the relation is a consideration for the promise made to him and the relation hath loss by not receiving the value of his father's labour but the law will not put a stranger in the place of the promisee, because there is no consideration why the payment shall be made to a stranger from whom the consideration did not rise.[45]

This laborious reasoning is not very convincing to the modern reader, and it is of interest in the present context for precisely that reason. Gilbert felt the necessity of formulating a proposition that would explain and justify the decision in *Dutton v Poole*, while maintaining a general rule of privity of contract and excluding the 'stranger.' Several possible propositions are suggested by this passage: the promise is enforceable where promisee and beneficiary are close relatives; the promise is enforceable where the promisee owes a moral duty to confer the benefit on the beneficiary; the promise is enforceable where some asset of the promisee that would otherwise have been transferred to the beneficiary is transferred instead to the promisor in exchange for the promise. All these formulations have difficulties, and it is not the object of the present inquiry to consider which is preferable, or which most accurately represented the law in 1678 or in Gilbert's own time. The significant point is that Gilbert felt the need to articulate a proposition, or principle, that would satisfactorily explain and justify *Dutton v Poole*, while at the same time stating the law as it was perceived to be when he wrote, and laying down a rule that would (in his and his readers' opinion) be satisfactory for the disposition of future cases.

The reason for the conclusion in *Dutton v Poole* plainly had something to do with general considerations of justice, including unjust enrichment. As the report puts it, 'the son hath the benefit by having of the wood, and

[44] S Comyn, *The Law of Contracts* (1826) vol 1, 27; Addison, above n 1, at 247–8, quoting Lord Mansfield in *Martyn v Hind* (1779) 1 Doug 142, 99 ER 94, 96: 'it is difficult to conceive . . . how a doubt could have been entertained in the case of *Dutton v Poole*'; Chitty, *A Treatise upon the Law of Contracts*, 2nd edn above n 23, at 48 (though with some reservation).

[45] Gilbert, above n 10, at n 82 (some punctuation added, abbreviations expanded, and capitalisation removed).

the daughter hath lost her portion by this means'.[46] Of course, the phrase 'unjust enrichment' was not in use in the seventeenth century, but plainer language could scarcely have been found to express the idea that the son had been unjustly enriched at the expense of the daughter. Gilbert also, in the passage just quoted, mentions that 'the son hath benefit by it', and makes the argument that the daughter suffers a loss by the transfer of the value of her father's 'labour', (that is, a valuable asset, in this case the timber), to the son instead of to her.

In the nineteenth century, *Dutton v Poole* came to seem inconsistent with the principle of consideration, and the case was rejected in *Tweddle v Atkinson*.[47] Crompton J expressly recognised that the law had changed:

> At the time when the cases which have been cited [these included *Dutton v Poole*] were decided the action of assumpsit was treated as an action of trespass upon the case, and therefore in the nature of a tort; and the law was not settled, as it is now, that natural love and affection is not a sufficient consideration for a promise upon which an action may be maintained; nor was it settled that the promisee cannot bring an action unless the consideration for the promise moved from him. The modern cases have, in effect, overruled the old decisions; they shew that the consideration must move from the party entitled to sue upon the contract. . . . I am prepared to overrule the old decisions, and to hold that, by reason of the principles which now govern the action of assumpsit, the present action is not maintainable.[48]

This is a very revealing passage, and shows the self-consciously novel view of principle that came to predominate in the mid-nineteenth century. The 'principles which now govern the action of assumpsit' required the overruling of 'the old decisions', even though no one doubted the justice (as between the parties) of *Dutton v Poole*, the case had stood for nearly 200 years, and the Queen's Bench had no actual power to overrule it.[49] Blackstone's treatment of assumpsit as part of private wrongs, less that 100 years old, was summarily consigned to a primordial period of unprincipled ignorance, which prevailed until the law was 'settled, as it is now'. These last words reflect the Victorian confidence in the perfectability, and nearly-attained perfection, of English social institutions, including law.

The approach in *Tweddle v Atkinson* was confirmed by the House of Lords in 1915. The concept of principle was again prominent. Viscount Haldane said that:

> in the law of England certain principles are fundamental. One is that only a person who is a party to a contract can sue on it. Our law knows nothing of a jus quaesitum tertio arising by way of contract. . . . A second principle is that if

[46] Above n 43, at 524.
[47] (1861) 1 B & S 393, 121 ER 762.
[48] *Ibid*, at 764.
[49] As pointed out by Blackburn J, *ibid*.

> a person with whom a contract not under seal has been made is to be able to enforce it consideration must have been given by him to the promisor or to some other person at the promisor's request. These two principles are not recognised in the same fashion by the jurisprudence of certain Continental countries or of Scotland, but here they are well established.[50]

The link between this conclusion and the doctrine of consideration was emphasised by Lord Dunedin, who, comparing Scots law, to its advantage, with English law, said pointedly:

> I confess that this case is to my mind apt to nip any budding affection which one might have had for the doctrine of consideration. For the effect of that doctrine in the present case is to make it possible for a person to snap his fingers at a bargain deliberately made, a bargain not in itself unfair, and which the person seeking to enforce it has a legitimate interest to enforce. Notwithstanding these considerations I cannot say that I ever had any doubt that the judgment of the Court of Appeal was right [as to English law].[51]

Lord Haldane's assertion that privity of contract was a 'fundamental principle' was challenged by Denning LJ, who said, in 1949,

> Counsel . . . says that the plaintiffs cannot sue. He says that there is no privity of contract between them and the board, and that it is a fundamental principle that no one can sue on a contract to which he is not a party. That argument can be met either by admitting the principle and saying that it does not apply to this case, or by disputing the principle itself. I make so bold as to dispute it. The principle is not nearly so fundamental as it is sometimes supposed to be. It did not become rooted in our law until the year 1861 . . . and reached its full growth in 1915.[52]

However, in 1962 the principle was reasserted by the House of Lords (Lord Denning, then himself a member of the House of Lords, dissenting). Viscount Simonds said

> Learned counsel . . . met [the argument for enforcement] . . . by asserting a principle which is, I suppose, as well established as any in our law, a 'fundamental' principle, as Lord Haldane called it, . . . an 'elementary' principle, as it has been called times without number, that only a person who is a party to a contract can sue upon it. 'Our law', said Lord Haldane, 'knows nothing of a jus quaesitum tertio arising by way of contract'. Learned counsel . . . claimed that this was the orthodox view and asked your Lordships to reject any proposition that impinged upon it. To that invitation I readily respond. For to me heterodoxy, or, as some might say, heresy, is not the more attractive because it is dignified by the name of reform. Nor will I easily be led by an undiscerning zeal for some abstract kind of justice to ignore our first duty, which is to administer

[50] *Dunlop Pneumatic Tyre Co Ltd v Selfridge & Co Ltd* [1915] 1 AC 847 (HL) 853.
[51] *Ibid*, at 855.
[52] *Smith v River Douglas Catchment Board* [1949] 2 KB 500 (CA) 514.

> justice according to law . . . The law is developed by the application of old principles to new circumstances. Therein lies its genius. Its reform by the abrogation of those principles is the task not of courts of law but of Parliament [two decisions of Lord Denning were cited here and rejected].[53]

This exceeds the forceful, and presses the limits of judicial courtesy. One could scarcely imagine a more powerful assertion of the absolute immutability of legal principle. Yet, surprising as this may seem, it was in effect abandoned within a few years, when the opposite result was reached by the Privy Council on similar facts in the New Zealand case of *The Eurymedon*.[54] Lord Wilberforce, giving the judgment of the majority of the Judicial Committee, said that the conclusion (enforcement by the third party) could 'be given within existing principles', but he had considerable difficulty in explaining the result in conventional terms. The Board was heavily influenced by considerations of commercial convenience and general considerations of fairness. Referring to an American case, Lord Wilberforce said that 'commercial considerations should have the same force on both sides of the Pacific', and said that he desired 'to give effect to the clear intentions of a commercial document' adding:

> It should not be overlooked that the effect of denying validity to the clause would be to encourage actions against servants, agents and independent contractors in order to get round exemptions (which are almost invariable and often compulsory) accepted by shippers against carriers, the existence, and presumed efficacy, of which is reflected in rates of freight. They see no attraction in this consequence.[55]

This was a clear victory for considerations of commercial convenience and general considerations of justice, and an effective abandonment of what had so recently been asserted by the House of Lords as 'fundamental principle'.

In a modern case having some parallels with the old case of *Dutton v Poole*, namely, *Beswick v Beswick*, an uncle transferred a coal business (his only substantial asset) to his nephew in exchange for the nephew's promise to pay an annuity to the uncle's widow.[56] The case resembles *Dutton v Poole* in that the promisor had actually benefited by receiving a valuable asset that would otherwise have benefited the plaintiff (because the uncle would otherwise have made other provision for her), differing in this respect from *Tweddle v Atkinson*, where no payments had been made. As Gilbert might aptly have said, the nephew had benefited by receiving the value of his uncle's labour, and the widow had lost correspondingly. But

[53] *Scruttons Ltd v Midland Silicones Ltd* [1962] 1 AC 446 (HL) 467–8.

[54] *New Zealand Shipping Co Ltd v A M Satterthwaite & Co Ltd* [1975] 1 AC 154 (PC) 169 [*The Eurymedon*].

[55] *Ibid*, at 169.

[56] [1968] AC 58 (HL).

this feature of the cases, of such obvious importance to every consideration of justice between the parties, was made to appear irrelevant by the principle established in *Tweddle v Atkinson* combined with an over-rigid scheme of categorisation that excluded considerations of unjust enrichment.[57] It is little to the credit of the law to establish, in the name of 'principle', rules that require the court to close its eyes to factors crucial to the attainment of justice.

The promise in *Beswick v Beswick* was held to be enforceable, but only because the widow happened to be the administratrix of the uncle's estate, and so entitled, in the opinion of the House of Lords, to a decree of specific performance. It is plain that the court was influenced by the general considerations of justice just mentioned. The law lords described the nephew's conduct as 'an unconscionable breach of faith',[58] and the possibility of there being no remedy as 'grossly unjust',[59] and 'repugnant to justice and [such as to] fulfil no other object than that of aiding the wrongdoer'.[60] The use of specific performance for this purpose was highly unusual. The remedy was given not, of course, because there was anything in the nature of money that could not compensate, but in order to circumvent the rule of privity, which Lord Pearce called 'a mechanical defect of our law'.[61] Lord Reid indicated, contrary to Viscount Simonds' view, that reform of the law would, if necessary, be within the proper power of the court.[62] Statutory reform of English law had been recommended by the Law Revision Committee in 1937, but nothing was done by Parliament until the enactment of the Contracts (Rights of Third Parties) Act, 1999.

In Canada, the English law of the late nineteenth and twentieth centuries was strictly followed, leading to a result in *Greenwood Shopping Plaza Ltd v Beattie*[63] which can, without much exaggeration, be called absurd.[64] The lessor of business premises had covenanted with its tenant to insure against fire. A loss by fire occurred, allegedly caused by the negligence of two of the tenant's employees, and the Supreme Court of Canada held that the lessor (and its insurer) was entitled to sue the two employees individually for the whole of the loss, even if the proper interpretation of the contract was that it had promised not to do so. The Supreme Court of Canada said

[57] See S Waddams, *Dimensions of Private Law: Categories and Concepts in Anglo-American Legal Reasoning* (Cambridge, Cambridge University Press, 2003).
[58] Lord Hodson, above n 56, at 83.
[59] Lord Reid, *ibid*, at 73.
[60] Lord Pearce, *ibid*, at 89.
[61] *Ibid*.
[62] *Ibid*, at 72. To similar effect Lord Scarman in *Woodar Investment Development Ltd v Wimpey Construction UK Ltd* [1980] 1 WLR 277 (HL) 300.
[63] [1980] 2 SCR 228, 111 DLR (3d) 257.
[64] See below n 80.

that 'the rule of privity . . . since *Tweddle v Atkinson* . . . has had decisive effect in this branch of the law. There are many cases which have applied this principle.'[65] As the Chief Justice of the Nova Scotia Court of Appeal had pointedly commented, this result '[flew] in the face of common sense, modern commercial practice and labour relations'.[66]

When the issue arose again in *London Drugs Ltd v Kuehne & Nagel International Ltd* the Supreme Court of Canada took a very different view.[67] The facts were rather similar to those in *Greenwood*. The plaintiff stored a valuable transformer with the defendant warehouser, agreeing to limit liability to $40. The transformer was damaged by the negligence of two employees, and, as in *Greenwood*, the owner sued the employees personally. The plaintiff's counsel relied on 'longstanding, established and fundamental principles of law', no doubt with some confidence of success in view of the quite recent decision of the court in *Greenwood*. However, Iacobucci J, giving the judgment of the majority of the court, decided in favour of the employees. Iacobucci J could easily have found that the case fell into one of the established exceptions to the doctrine of privity, but he chose instead to deal with the issue directly, saying 'I prefer to deal head-on with the doctrine of privity and to relax its ambit in the circumstances of this case'.[68] He considered that the strict rule should be relaxed for reasons of 'commercial reality and common sense'.[69] Similar expressions were repeated: 'sound commercial practice and justice',[70] 'the reasonable expectations of all the parties to the transaction',[71] 'the underlying concerns of commercial reality and justice',[72] 'commercial reality',[73] a result that made 'sense in the modern world',[74] 'sound policy reasons',[75] 'commercial reality and justice',[76] and 'modern notions of commercial reality and justice',[77] ideas that were contrasted, to their advantage, with 'a strict application of the doctrine of privity',[78] and 'the rigid retention of a doctrine that has undergone systematic and substantial attack'.[79] He said that it would be

[65] Above n 63, at 263.
[66] *Greenwood Shopping Plaza Ltd v Neil J Buchanan Ltd* (1979) 99 DLR (3d) 289 (NSAD) 295 (MacKeigan CJNS).
[67] [1992] 3 SCR 299, 97 DLR (4th) 261.
[68] *Ibid*, at 341 (DLR).
[69] *Ibid*, at 342.
[70] *Ibid*, at 348.
[71] *Ibid*.
[72] *Ibid*.
[73] *Ibid*, at 360.
[74] *Ibid*, at 364.
[75] *Ibid*.
[76] *Ibid*, at 365.
[77] *Ibid*, at 370.
[78] *Ibid*, at 361.
[79] *Ibid*, at 358.

'absurd in the circumstances of this case to let the appellant go around the limitation of liability clause by suing the respondent employees in tort'.[80]

The court did not, however, simply abolish the doctrine of privity, but created a limited exception, which Iacobucci J described as an 'incremental change'.[81] Many commentators and courts had said, as we have seen, that it was a 'principle'—often indeed called a 'fundamental principle'—of English and Canadian law that only a party to a contract could sue on it. Iacobucci J himself described privity of contract as 'an established principle in the law of contracts [which] should not be discarded lightly'.[82] This phrase demonstrates the elusive meaning of the idea of principle. Evidently a 'principle,' even though 'established', is not necessarily determinative of legal issues, but can be 'discarded' for sufficient reason (though not 'lightly'). Although in that sentence Iacobucci J called privity of contract a 'principle', more often he described privity as a 'doctrine' and the *exception* to it as 'principled'.[83] Some might wish that the court had made a more radical change. On this it may be remarked first that an exception on such general grounds as 'commercial reality and justice' *is* in reality a very far-reaching change, and second that there are good reasons for caution. A simple declaration by the court that the rule of privity was abolished would have *compelled* future courts to enforce contracts for the benefit of third parties; the recognition of a limited exception, on the other hand, has the effect of *empowering* future courts in appropriate cases to enforce such contracts, but not compelling them to do so. There are good reasons not to lay down a rule that all contracts for the benefit of third parties must always be enforced. There are two particularly difficult cases. One is the case of the incidental beneficiary—one who would have benefited by performance of the contract but not a person on whom the contracting parties intended to confer rights. The other difficult case is where the original contracting parties seek to rescind or modify the contract. Sometimes it is appropriate for them to do so, and sometimes it is more appropriate to require the consent of the third party, but it is not easy to formulate a universal rule on the point. The effect of the decision in *London Drugs* was to reintroduce flexibility to the common law, and to enable the lower courts to reach fair and just results. This approach has advantages over statutory reform, which, as the English experience shows, is apt to lead to unexpected anomalies and complexities.[84] In the Canadian

[80] *Ibid*, at 363.
[81] *Ibid*, at 366.
[82] *Ibid*, at 358.
[83] See below n 86.
[84] The compexities of the English statute are forcefully demonstrated by R Stevens, 'The Contracts (Rights of Third Parties) Act, 1999' (2004) 120 *LQR* 292.

context there is the added point that uniform provincial legislation on the matter could not realistically have been anticipated.

In the subsequent case of *Fraser River Pile & Dredge Ltd v Can-Dive Services Ltd* Iacobucci J, giving the judgment of the whole court, extended the *London Drugs* case to a case involving waiver by an insurer of subrogation rights.[85] The decision shows that the recognition of third party rights in contracts is not limited to any particular class of contract, and there seems no reason why third party rights should not be recognised in any case where considerations of justice require it. Iacobucci J described the *London Drugs* case as having introduced a 'principled exception to the common law doctrine of privity of contract'.[86] Principle and policy were closely associated in his mind. In *Fraser River* he speaks of 'policy reasons in favour of an exception',[87] and, as in *London Drugs* of 'common sense and commercial reality'.[88] Sometimes a distinction has been made between principle and policy, but it is evident that in Iacobucci J's mind they were not opposed: on the contrary, good policy was an essential aspect of sound principle. The same could be said of many influential judges and commentators in the nineteenth and twentieth centuries.

Perhaps the greatest difficulty posed by the English doctrine of consideration has been the treatment of reliance. Subsequent reliance on a promise does not constitute consideration, but the consequence of refusing enforcement can be severe injustice. One of the simplest imaginable examples of reliance arises when a landowner promises to give the land to another person (for example, a relative) and the other person, relying on the promise, builds on the land. The promisor (or, as it has more usually been, his or her estate after death) then seeks to revoke the promise. These facts have presented a problem for Anglo-American law because the transaction, being gratuitous, is not enforceable as a contract. Property has not been legally transferred; neither is the promisor guilty of any tort. Nevertheless the courts of equity gave a remedy to the promisee.[89] These cases could not be reconciled with orthodox contract doctrine respecting consideration and have therefore been ignored or marginalised by many writers on contract law. They have usually been described by commentators as cases of proprietary estoppel but this phrase is scarcely explanatory. In some of the cases avoidance of unjust enrichment (though not always by

[85] [1999] 3 SCR 108, 176 DLR (4th) 257.
[86] *Ibid*, at [24].
[87] *Ibid*, at [40].
[88] *Ibid*, at [25].
[89] *Dillwyn v Llewelyn* (1862) 4 De G F & J 517, 45 ER 1285 (CA); *Ramsden v Dyson* (1866) LR 1 HL 129; *Wilmott v Barber* (1880) 15 Ch D 96; *Plimmer v Wellington* (1884) 9 App Cas 699 (PC); *Inwards v Baker* [1965] 2 QB 29 (CA); *Crabb v Arun DC* [1976] Ch 179 (CA); *Greasley v Cooke* [1980] 1 WLR 1306 (CA); *Stiles v Tod Mountain Devt Ltd* (1992) 64 BCLR (2d) 366 (SC).

that name) was evidently a crucial factor: in the leading case of *Dillwyn v Llewelyn*, for example, the plaintiff had expended the very large sum of £20,000 in improving land that was originally worth only £1,500.[90] One of the considerations in the mind of a court faced with such facts has been the enrichment that would enure to the defendant if no measure of enforcement were available, and one of the reasons for favouring proprietary estoppel as a rule is that unjust enrichment is *very apt* to occur in such circumstances, and so the rule tends to prevent unjust enrichment. But unjust enrichment has not been present in every particular case,[91] and the remedy has not normally been measured by enrichment. As Peter Birks wrote, 'the doctrine … has a *dimension* to it which has nothing to do with restitution/unjust enrichment'.[92] This is true, but it does not follow that considerations of unjust enrichment have been irrelevant. Many of the cases have had the effect of protecting reliance, but where the plaintiff becomes effectively the owner of the land the measure of recovery is not restricted to out of pocket loss. Non-contractual reliance has sometimes been protected by concepts of wrongdoing, but in these cases there is no wrongdoing in the ordinary sense. Though the phrase 'equitable fraud' has sometimes been employed, no actual proof of wrongdoing has been required: the defendant acts fraudulently, in the eyes of equity, by failing to do what is just. As expressed in *Willmott v Barber* (1880), 'the plaintiff must prove that he [the defendant] has acted fraudulently, *or* that there has been such an acquiescence on his part as would make it fraudulent for him *now* to assert his legal rights'.[93]

The word 'fraudulent' in the last clause of this passage means 'unjust', and cannot be explained except in terms of concepts other than wrongdoing. The defendant must, by action or inaction, induce the plaintiff's reliance, but no proof of intention to mislead or deceive is required.[94] The only 'fraud' required to be proved is an unwillingness to do what equity considers just. A similar comment may be made in relation to the concept of unconscionability.[95] If equity protects the plaintiff's reliance, it will be,

[90] *Dillwyn v Llewelyn* (1862) 6 LT 878 (CA) 879.

[91] See J Beatson, *Anson's Law of Contract*, 27th edn (Oxford, Oxford University Press, 1998) 119.

[92] P Birks, *Introduction to the Law of Restitution* (Oxford, Clarendon Press, 1985) 290 (emphasis added). R Goff and G Jones, *The Law of Restitution* (London, Sweet & Maxwell, 1965) and P Maddaugh and J McCamus, *The Law of Restitution* (Toronto, Canada Law Book, 1990) included discussion of these cases in their books, but A Burrows, *The Law of Restitution* (London, Butterworths, 1993) wrote flatly at 404 that these cases 'have not been restitutionary'.

[93] *Willmott v Barber* (1880) 15 Ch D 96, 106 (emphasis added); *Gerrard v O'Reilly* (1843) 3 D & War 414 (Ir Ch).

[94] *Willmott v Barber*, *ibid*, at 105 (Fry J). This is a common usage in equity. See L Sheridan, *Fraud in Equity* (London, Sir Isaac Pitman and Sons, 1957).

[95] See M Spence, *Protecting Reliance: the Emergent Doctrine of Equitable Estoppel* (Oxford, Hart Publishing, 1999) 55–66; *Giumelli v Giumelli* (1999) 161 ALR 473 (HCA).

by that very fact, against conscience for the defendant to defeat it. 'Inequitable', 'fraudulent', 'unconscionable' and 'unconscientious' have been, in this context, four ways of saying the same thing.[96]

Where a party to a contract indicates that strict rights will not be enforced, equity has in some cases prevented that party from resuming the strict rights where it would be inequitable to do so. In *Hughes v Metropolitan Railway* a landlord who was entitled to demand repairs on six months notice, gave the notice, but then indicated by conduct that the running of the period of notice was suspended. The House of Lords held that the landlord could not enforce its strict rights without giving the tenant a reasonable opportunity of compliance. 'Principle' might be thought to permit the landlord to assert its strict rights, because there was no consideration for the implied promise to suspend them, but the court relied on another proposition, also called a principle—indeed actually called 'the first principle':

> [I]t is the first principle upon which all Courts of Equity proceed, that if parties who have entered into definite and distinct terms involving certain legal results—certain penalties or legal forfeiture—afterwards by their own act or with their own consent enter upon a course of negotiation which has the effect of leading one of the parties to suppose that the strict rights arising under the contract will not be enforced, or will be kept in suspense, or held in abeyance, the person who otherwise might have enforced those rights will not be allowed to enforce them where it would be inequitable having regard to the dealings which have thus taken place between the parties.[97]

For 'the first principle', this is not very compendiously stated, and the reader might wonder how many other such 'first principles' there might be, whether they are a closed number, and whether they can all be definitively formulated.

In *Central London Property Trust Ltd v High Trees House Ltd*, a landlord had reduced the rent of an apartment building during wartime, and the question arose whether, after the war, the landlord could resume its right to the full rent. Although the landlord made no claim for arrears, the judge, Denning J, took the opportunity to say that arrears could not have been claimed. Denning J formulated the following proposition:

> I prefer to apply the principle that a promise intended to be binding, intended to be acted on and in fact acted on, is binding so far as its terms properly apply.[98]

This 'principle' had nowhere previously been formulated, and a few years later Denning himself (then a member of the Court of Appeal) was forced

[96] See A Robertson, 'Reasonable Reliance in Estoppel by Conduct' (2000) *University of New South Wales Law Review* 87, 96–7.

[97] (1877) 2 App Cas 439 (HL) 448.

[98] [1947] 1 KB 130, 136.

to retreat substantially from it.[99] The result of this advance and hasty retreat has been a very high degree of uncertainty about the scope of the alleged principle. Denning LJ used the word 'principle' five times in his fairly short judgment,[100] but it is not at all clear what principle the decision applied. Unanswered questions include whether a previous legal relationship between the parties is necessary, whether detrimental reliance is essential, whether estoppel can create new legal rights, whether the promisee's remedy is limited to protection of reliance, and whether strict rights can be resumed on reasonable notice. Denning tried again in 1982 to formulate a 'general principle' that would account for all instances of estoppel:

> The doctrine of estoppel is one of the most flexible and useful in the armoury of the law. But it has become overloaded with cases. . . . It has evolved during the last 150 years in a sequence of separate developments: proprietary estoppel, estoppel by representation of fact, estoppel by acquiescence, and promissory estoppel. At the same time it has been sought to be limited by a series of maxims: estoppel is only a rule of evidence, estoppel cannot give rise to a cause of action, estoppel cannot do away with the need for consideration, and so forth. All these can now be seen to merge into one general principle shorn of limitations. When the parties to a transaction proceed on the basis of an underlying assumption—either of fact or of law—whether due to misrepresentation or mistake makes no difference—on which they have conducted the dealings between them—neither of them will be allowed to go back on that assumption when it would be unfair or unjust to allow him to do so. If one of them does seek to go back on it, the courts will give the other such remedy as the equity of the case demands.[101]

But neither can this formulation be adjudged a success. In 1996 Millett LJ said that the 'attempt to demonstrate that all estoppels . . . are all governed by the same requirements has never won general acceptance'.[102] In 2001 Lord Goff said, of the statement of Lord Denning's just quoted, 'This broad statement of the law is most appealing. I yield to nobody in my admiration for Lord Denning; but it has to be said that his attempt in this passage to identify a common criterion for the existence of various forms of estoppel . . . is characteristically bold.'[103] Lord Goff then, having quoted a statement from a treatise to the effect that estoppel should not be permitted to undermine the doctrine of consideration, added:

> I myself suspect that this statement may be too categorical; but we cannot ignore the fact that it embodies a fundamental principle of our law of contract. The

[99] *Combe v Combe* [1951] 2 KB 215 (CA).
[100] *Ibid*, at 219–20.
[101] *Amalgamated Investment and Property Co Ltd v Texas Commerce International Bank Ltd* [1982] QB 84 (CA) 122.
[102] *First National Bank v Thomson* [1996] Ch 231 (CA) 236.
[103] *Johnson v Gore Wood & Co* [2002] 2 AC 1 (HL) 39–40.

> doctrine of consideration may not be very popular nowadays; but although its progeny, the doctrine of privity, has recently been abolished by statute, the doctrine of consideration still exists as part of our law.[104]

Lord Goff then said, of estoppel, 'in the end I am inclined to think that the many circumstances capable of giving rise to an estoppel cannot be accommodated within a single formula, and that it is unconscionability which provides the link between them'.[105] By this he meant that the underlying reason for estoppel is that it would be unfair for a person, having made an assertion that induces another to act to her detriment, to go back on the assertion. This approach is capable, in effect, of enforcing promises, as is illustrated by the Australian case of *Waltons Stores (Interstate) Ltd v Maher*.[106] There an owner of land demolished a building and commenced construction of a new one to the specifications of a prospective tenant. No binding lease was ever effected, but the prospective tenant was held to be estopped (in the view of the majority of the court) from retreating from an implied promise to complete the contract. The effect was that, although there was no contract, the prospective tenant was bound by precisely those obligations that would have existed if there had been a contract, a state of affairs that might not unreasonably be summarised by saying that, in effect, there *was* a contract.

'Principle' means beginning (*principium*), but it is evident that this is not a point in historical time: it is more akin to a conceptual or notional origin, root or source. From this it follows that principles are liable to vary according to the concepts or notions of the speaker or writer who uses the word. The search for principle in law is not a purely historical inquiry, but it has a historical dimension. The history of the law includes the history of failed attempts to formulate principles, of which we have seen several instances in the present context, and the questions of what concepts have been employed in the past, and how widely they have been held, are historical questions. Addison said that contract law was founded on principles that were immutable, eternal and universal, but the present inquiry suggests that, in respect of consideration, it has not been possible to formulate principles that have been stable even over a short period of time in the history of one legal system. 'Principle' has, to say the least, been a very versatile tool, changing its shape to suit diverse purposes: in the hands of Viscount Simonds, principle prohibited reform; in the hands of Crompton J, and of Lord Denning, principle demanded reform—but in entirely opposite directions. It does not follow that the concept of principle has been useless or unimportant; on the contrary, the constant appeal by

[104] *Ibid*, at 40.
[105] *Ibid*, at 41.
[106] (1988) 164 CLR 387 (HCA).

courts and writers to principle suggests that the *concept* of principle, though not the formulation of any particular proposition, has been perceived as essential to legal reasoning. A principle is a proposition that links past, present and future: a proposition, to be accepted as a legal principle, must account satisfactorily for past decisions thought to have been rightly decided; it must resolve the current dispute in a manner perceived to be satisfactory; and it must be capable of resolving in a satisfactory manner all future instances that can be envisaged. The appeal to principle implies that propositions, even when it is known that they are being formulated and applied for the first time, can be traced to some sort of conceptual or notional beginning, and this implication has been an important part of what has enabled an uncodified system to combine flexibility and the ability of the law to change, with the preservation of continuity, stability, and coherence. Paradoxically, the concept of principle has succeeded only by appearing to be what it is not. At each point of time principles have appeared to be stable, as they must appear in order to be called principles, but from a historical perspective it can be seen that they have been constantly susceptible to change.

4

Contractual Interpretation at Common Law and Civil Law: An Exercise in Comparative Legal Rhetoric

CATHERINE VALCKE

THAT CIVILIANS AND common lawyers approach contractual interpretation differently is well established.[1] Also well established is that, as with most issues of private law, the difference lies not so much in the outcomes of the decided cases as in the means that are used to reach these outcomes.[2] That is to say, judges in the two systems reach outcomes that are on the whole very similar, but they do so by deploying very different rules and institutions. It is well-known, for example, that civil law judges often appeal to some general notion of good faith to arrive at conclusions which common law judges instead reach by invoking implied terms.[3]

Now the particular reasons offered (or not offered) in support of the choice of particular rules and institutions in each system are arguably far more significant than the rules and institutions themselves. Because the weight and legitimacy given to legal decisions are largely determined by the

[1] B Nicholas, *The French Law of Contract* (Oxford, Clarendon Press, 1992) 47ff; R David and D Pugsley, *Les contrats en droit anglais* (Paris, LGDJ, 1985) 251ff.

[2] A Burrows and E Peel, 'Overview' in A Burrows and E Peel (eds), *Contract Terms* (Oxford, Oxford University Press, 2007) 3, 7–8; S Vogenauer, 'Interpretation of Contracts: Concluding Comparative Observations' in Burrows and Peel, *ibid*, at 123, 149–50. For one recent Supreme Court of Canada decision where this difference might have nonetheless determined the outcome, see: *Double N Earthmovers Ltd v City of Edmonton* [2007] 275 DLR (4th) 577 [*Double N Earthmovers*]. The case turned on the interpretation to be given to two implied terms contained in a tendering contract. The five judges in the majority, three of whom were civilian (Lebel, Deschamps, and Fish JJ), favoured a narrow reading of the terms, which resulted in the tenderer not being found in breach; the four dissenting judges' broader reading would have resulted in a finding of breach. Had any one of the three civilian judges endorsed the broader reading, judgment would have been rendered against the tenderer.

[3] Nicholas, above n 1, at 49–50.

persuasiveness of the reasons offered in support of these decisions, legal actors intent on defending a particular decision can be expected to appeal to the reasons that are considered most persuasive in their community. And the reasons considered most persuasive in any legal community naturally are those that most directly tap into its core values, namely, its animating conceptions of law, liberty, equality, the role of the State, and so on. The reasons offered in support of legal decisions in any given community thus can be taken to mirror that community's fundamental legal values. What is more, the bare fact that the legal actors appealing to these reasons *perceive* them to be persuasive in their community is itself significant: being themselves members of that community, their perceptions as to the community's values to a certain extent are constitutive of those values.[4] Verging on circularity, therefore, the reasons offered in support of legal decisions are significant both because they reflect the community's core legal values and because the very fact that they are being heralded as reflective of these values makes them so.[5] As the embodiment then of a legal system's core values, the reasoning offered in support of particular choices of rules and institutions is clearly useful for the purpose of coming to an internal understanding of that system taken as a whole. But it also is useful for the purpose of understanding the rules and institutions themselves. For instance, the reasons that common law judges offer for interpreting contracts through implied terms say much about the fundamental values of the common law system as a whole, as well as about the English doctrine of implied terms. And then similarly the reasons given by civilians in support of their judges resorting to good faith say much about the fundamental values underlying the civil law, as well as about the nature of the civilian notion of good faith. For to 'understand' a legal rule is to be able to locate it within the larger edifice of the legal system to which it belongs, to connect it with the other rules of that system; such a connection can only be established through the values reflected in the reasons given. A meaningful comparative understanding of the various rules and institutions deployed in each different system for the purpose of interpreting contracts therefore will necessarily involve comparing these rules and institutions *as they relate to one another and to the other elements of their respective systems through the values embodied in the attendant reasoning*.

The following account of contractual interpretation at common law and civil law focuses accordingly on the reasoning used, that is, on the

[4] This is not to say that the legal values of a community are *nothing but* a construction of that community. It could well be that this layer of community-specific values is superimposed upon or somewhat interwoven with another, deeper layer of universal values which in contrast would by definition be common to all legal systems.

[5] See generally N Luhmann, 'L'unité du système juridique' (1986) 31 *Archives de philosophie du droit* 163.

arguments that are offered, to justify contractual interpretation, and much less upon the various rules and institutions of those systems. In that sense, this account qualifies as an exercise in comparative legal rhetoric. Its aim is this: through the reconstruction of the respective conceptions of the contractual enforcement process embodied in the arguments of each system, the article seeks to connect the rules and institutions of each system to one another and to contrast them between systems. These two tasks are accomplished in tandem, not sequentially: as the reconstruction process unfolds and the rules on each side are being connected, the contrast between the systems becomes increasingly apparent. Part I describes and justifies the parameters of this exercise. No comparative undertaking can proceed without first identifying an appropriate neutral common basis, or *tertium comparationis,* upon which to conduct the comparison.[6] A common language, conceptual territory, and set of criteria must be established which are sufficiently abstract to apply to the two terms under comparison without distorting the identity of either, yet not so abstract as to be meaningless. Part I aims accordingly to expose the neutrality of our working definition of 'contractual interpretation,' of the process by which the pool of relevant legal materials was delineated, and of the criteria used to conduct the actual comparison. The comparison takes place in Parts II and III, which are each devoted to surveying civil law and common law materials on contractual interpretation. For reasons of convenience, the discussion of Part II is limited to French and Quebec legal materials,[7] and that of Part III is limited to English and Canadian materials.

I. PRELIMINARIES

The legal materials targeted by the present study are those pertaining to 'contractual interpretation' writ large, that is, to contractual interpretation understood as the task of determining the normative content of a contract. This task is distinguished from those of determining whether a contract exists in the first place, whether it was in fact breached, and if so, what is to be done to redress this breach. We will see that in both systems the term 'interpretation' has been given a variety of more specific and at times divergent meanings. Precisely for that reason it is necessary to adopt, at

[6] Since Radbruch (*über die Methode der Rechtsvergleichung*, MKSR II, 423, 1905/06), the notion of *tertium comparationis* is a staple of comparative law literature. See in particular: H Kötz, 'Comparative Law in Germany Today' [1999] *Revue internationale de droit comparé* 753, 758ff.

[7] A civil law survey that does not include German legal materials is obviously incomplete, but unlike Quebec materials, German materials are sufficiently different from French materials to warrant a separate treatment. See eg, Vogenauer, above n 2, who runs the French and German analyses in parallel.

least provisionally, as broad a definition as possible, one which captures all that is *seen* as pertaining to contractual interpretation in either system. For, as indicated, a study of legal rhetoric very much is an exercise in legal *self*-perception, an exercise in perception from the standpoint of the legal actors themselves.[8]

Among the materials on contractual interpretation so understood, special attention is given to those that feature legal actors choosing to support their decisions with some particular arguments *rather than others* also reasonably available to them, and especially to materials that show legal actors choosing, from the pool of available arguments, those that seem *least plausible*, all else being equal. For in the extreme case where only one argument is reasonably available, the legal actor has no choice. The fact that that argument ends up being used only reflects the absence of an alternative—it says nothing as to the actor's assessment of the argument's relative persuasiveness. In other words, persuasiveness being a matter of degree, judgments as to the persuasiveness of arguments are necessarily comparative: they cannot but proceed from the observation of cases where the legal actor chooses one argument or set of arguments *over* at least one available alternative. For example, in a contract case where the judge decides in favour of the defendant on the ground that 'the contract never formed,' the fact that the judge cites this reason is far more significant, for the purpose of assessing this reason's persuasiveness, if the facts are such that the alternative reason 'the contract is unenforceable on ground of unfairness' is also reasonably available. For only then can it reasonably be contended that the judge might have cited the reason 'the contract never formed' because the judge considered it more persuasive than the alternative. Such a comparative judgment would simply not be possible were the facts such that the reason 'the contract never formed' was the only one reasonably available to the judge. This example also underscores how the significance of any given choice of reasons, among a plurality, is inversely related to the plausibility of the reason chosen. The likelihood that judges chose 'the contract never formed' *because* they considered that reason more persuasive is greater where, all else being equal, that reason is, in the circumstances, less plausible than the alternative. For it then is possible to conclude that that reason was chosen *in spite* of its being less plausible than the alternative, whereas one would normally expect judges to choose

[8] The legal actors at play here are idealised legal actors in the same way that Dworkin's Hercules is an idealised judge (R Dworkin, *Taking Rights Seriously* (Cambridge, Harvard University Press, 1977) 184ff): they are infinitely intelligent and knowledgeable of the legal materials that make up their respective systems, they take these materials seriously, as the sole embodiment of legal authority in their system, and they view themselves as charged with the task of reconstructing these materials into a coherent conceptual entity. It is in that sense that exercises in legal rhetoric differ from exercises in, say, legal anthropology or legal sociology, which in contrast aim to study *actual* legal actors.

the most plausible among the available reasons. In sum, the less plausible a chosen reason, the more significant its choice becomes, and the more likely it is that that choice reflects a judgment about the relative persuasiveness of the various available reasons (and hence also the values peculiar to the legal system). The survey of Parts II and III accordingly focuses, to the extent possible, upon materials that show legal actors choosing a particular argument or set of arguments over other available ones, with specific attention being paid to cases where the chosen arguments were less plausible than their alternative.

The 'legal actors' under observation in Parts II and III mainly are the judges, on the English side, and *la doctrine* (the scholars), on the French side. While the decisions of all legal actors, not just those of judges or scholars, admittedly are relevant for the purpose of assessing what arguments are considered persuasive in the system,[9] there are good reasons here to focus on this particular set of legal actors. As a matter of convenience, English judges and French scholars remain the richest sources of easily accessible reasons in their respective systems. As a matter of principle, moreover, the tighter focus upon the legal actors considered formal or quasi-formal lawmakers in their respective systems is desirable for the purpose of keeping the impartiality of the comparison in check: it ensures that the comparison is directed at dimensions of the two systems that can reasonably be considered counterparts of one another.[10] And it indeed can reasonably be asserted that French scholars for present purposes are to French law what English judges are to English law. French judicial decisions are notoriously short on reasons and therefore it has traditionally fallen to the scholars to supply them.[11]

As for the neutral criteria with which to conduct the comparison, they are to be found in the common functions discharged by contract law across

[9] The arguments provided by legislators in support of their legislative decisions, by private parties in the context of contractual negotiations, by law students in support of their examination answers etc, can similarly be expected to be those considered most persuasive in the community, and thus reflective of the community's core values. See C Valcke, '"Precedent" and "Legal System" in Comparative Law: A Canadian Perspective' in E Hondius (ed), *Precedent and the Law* (Brussels, Bruylant, 2007) 85.

[10] Radical legal cultural relativists would deny the very possibility of establishing such 'counterparts' between legal systems, but their argument, pressed to its logical conclusion, would also rule out the very possibility of any form of inter-system comparison, despite some of them having provided highly illuminating comparisons. See eg, P Legrand, *Le droit comparé* (Paris, PUF, 1999).

[11] This explains the great authority ascribed to scholars at French law—certainly as compared to scholars at English law, but also as compared to judges at French law—similar to that enjoyed by the great jurists of Roman law. See generally RC Van Caenegem, *Judges, Legislators and Professors—Chapters in European Legal History* (Cambridge, Cambridge University Press, 1987) 67–111.

legal systems.[12] If only intuitively, it indeed is possible to think of contract law in all systems as serving at least two different and somewhat contradictory functions. The first is to 'enforce' contracts: strictly speaking, that is to bring about whatever it is that private parties want to be brought about. This first function can be described as 'consecrating', in so far as it entails doing no more than supplying the parties' original intention[13] with legal force, to elevate that intention from the realm of fact into the realm of law. The parties' own original conception of their respective obligations here is adopted (and consecrated), at least aspirationally, *as is*; no attempt is made to impose on the parties anything that they did not originally intend. The second function, which I call 'disciplining,' instead involves running *counter* to the parties' original intention. Rather than giving (legal) voice to the parties, the aim here is to dictate what the parties are allowed to intend, to substitute an acceptable legal intention for the parties' own. In that sense, the disciplining function is prescriptive, unlike the consecrating function, which aspires to remaining normatively neutral. The consecrating and disciplining functions are conceptually contradictory in so far as the first involves moving along with the parties, whereas the second entails pressing back against them.[14]

In practice, be it civil law or common law, the consecrating and disciplining functions of contract law cannot be so neatly distinguished, because both these functions inevitably are simultaneously at play in the enforcement of actual contracts. The contractual intention being enforced indeed always is the parties' original intention *as it has been processed by the court*: the parties provide the raw data which the court then proceeds to re-formulate, re-cast, transform into something deserving of the law's

[12] The focus upon the common functions discharged by the different legal rules and institutions in the different legal systems as the starting point for comparative analysis is known in comparative law literature as 'functionalism'. Although functionalism as a theory of comparative law has been much criticised (L-J Constantinesco, *Traité de droit comparé* (Paris, LGDJ, 1983) vol III, 63–71) its value as one of several methodological tools available to the comparatist is widely acknowledged: J Reitz, 'How to do Comparative Law' (1998) 46 *American Journal of Comparative Law* 617, 620–23; G Samuel, 'Epistemology and Comparative Law: Contributions from the Sciences and Social Sciences' in M Van Hoecke (ed), *Epistemology and Methodology of Comparative Law* (Oxford, Hart Publishing, 2004) 35, 38ff.

[13] It here is assumed that one such common intention exists. This assumption will be relaxed below.

[14] Although the consecrating/disciplining distinction obviously parallels the continental distinction between 'subjective' and 'objective' right (Droit/Rechts), these terms here are avoided both because they resonate differently at French and English law and because the continental distinction carries much baggage that here is superfluous and thus potentially more confusing than useful. The present use of the consecrating/disciplining distinction as *tertium comparationis* is justified, despite this distinction being connected with the continental one, in so far as, in this new incarnation, and with the caveat just mentioned, it is arguably sufficiently abstract to be usefully applicable to both systems without distorting either.

imprimatur. The final product therefore always is the result of a combination of consecrating and disciplining.[15] As indicated, however, it is not the final product of the contractual enforcement process in the two systems that interests us here, but rather how that process is viewed by the legal actors in these systems. And the views of the legal actors indeed appear to diverge in this regard, as I propose to show next. The view that emerges from the civil law materials is that of a process in which the consecrating and disciplining functions are neatly delineated from one another, and the disciplining function is by far the more prominent (Part II). The view that emerges from the common law materials in contrast is that of a process in which the two functions are merged, or more specifically, in which the disciplining function has been merged into the consecrating function in an apparent attempt to downplay the former's relative significance (Part III).

II. (FRENCH) CIVIL LAW MATERIALS ON CONTRACTUAL INTERPRETATION

The prominence of the disciplining function in the French conception of contract law is obvious from the outset, from the way that the contractual enforcement process is structured. As one must have determined that a contract exists before undertaking to determine its content, contract adjudication at French law, just like at English law (and any other minimally logically-inclined law), proceeds in that order. Also like at English law, moreover, contract formation is considered an issue to be objectively determined by the law makers, that is, independently from the parties' own views on the matter.[16] That is to say, French and English legal actors alike consider that it falls to the law makers, not the parties, to fix the conditions of existence of a contract and to determine whether these conditions are satisfied in any given case.[17] Inasmuch as these conditions, and the determination as to their satisfaction, can accordingly be said to be 'imposed' on the parties, it can be said that the issue of contract formation in both systems is treated as a 'disciplining' rather than a 'consecrating'

[15] Even in cases where the court apparently ends up enforcing the parties' intention as is, the disciplining function is present since the court then implicitly signifies its normative approval of that intention.

[16] The fact that consent—perhaps the most fundamental of the constitutive elements of the contract—itself is, as we will see, largely determined subjectively does not change the fact that the qualification of consent as such an element itself is the result of an objective determination.

[17] In so far as the 'gentlemen's agreements' cases of English law are seen as cases in which the court refuses enforcement because *the parties* declared that their agreement ought not be considered a binding contract, these cases would constitute an exception. See, eg, *Rose and Frank Company v JR Crompton & Brothers Ltd* [1923] 2 KB 261 (CA). But other readings of these cases are possible which justify non-enforcement on the basis that one or several of the (objectively determined) elements of a contract in fact are missing in these cases.

issue under the above definition. The two systems, however, differ in that, at French law (unlike at English law) a determination of the issue of whether a given contract was formed also determines in large part the distinct and subsequent issue of what that contract contains, with the result that the latter issue is, to a large extent, also determined objectively.

The existence of a *convention* at French law indeed is determined by the presence of one more elements than those familiar to the English lawyer. French law requires, in addition to an *échange de consentements* (offer and acceptance), the *capacité de contracter* (capacity to contract) of the parties, and a *cause* (roughly, consideration[18]), that there be 'a determinate *objet* (object) forming the subject matter of the agreement'.[19] In any contract dispute, therefore, the 'subject matter of the agreement'—also known as the 'juridical operation envisaged by the parties'[20]—is to be objectively established at the outset, as one of the conditions for the existence of the *convention*. But this identification of the *objet* for the purpose of establishing the existence of the *convention* also serves to determine much of the content of that *convention*, for it allows for both an early check on its *licéité* (legality) and its initial *classification*. Unlike at English law, where the control of the contract's legality cannot but intervene after the existence of the contract has been established,[21] the *objet* as a constitutive element of a *convention* at French law allows for one such control to occur at the initial stage of formation: contracts with illicit subject matters are deemed non-existent from the outset.[22] In addition, the initial identification of the *objet* allows for the *classification* of the *convention* under one of many

[18] It is well known that the French *cause* differs substantially from the English notion of consideration. Whereas the notion of consideration entails the mutuality of a bargain, the French *cause* has been interpreted broadly, so as to be deemed present even in gift transactions, in the form of 'donative intent'. French jurists have debated the issue *ad nauseam*, some of whom (the 'anti-causalists') suggesting that such a broad understanding of *cause* is tantamount to none at all, and that the whole notion therefore might as well be abandoned. See eg, G Ripert and J Boulanger, *Traité de droit civil* (Paris, LGDJ, 1956–59) vol 2, § 287.

[19] '*Un objet certain qui forme la matière de l'engagement*.' Art 1108 CC. (My translation—all English translations of French texts are mine unless otherwise indicated.) Art 1371 of the CCQ similarly provides that '[i]t is of the essence of an obligation that there be persons between whom it exists, a prestation which forms its object, and, in the case of an obligation arising out of a juridical act, a cause which justifies its existence'. (The CCQ, unlike the Code Napoléon (CN), has an official English version.)

[20] Art 1412 CCQ; F Terré, P Simler and Y Lequette, *Droit civil—Les Obligations*, 5th edn (Paris, Dalloz, 1993) §§ 257, 287.

[21] Where the court concludes from its examination of the content of the contract that this contract is illegal or against public policy, the contract is declared to be *retroactively* void *ab initio*. See *Archbolds (Freightage) Ltd v Spanglett Ltd* [1961] 1 QB 374 (CA).

[22] Art 1413 CCQ provides that '[a] contract whose object is prohibited by law or contrary to public order is null'. At French law, in contrast, the *licéité* requirement of the *objet* was articulated by *la doctrine* (see, eg, Terré *et al*, above n 20, at § 304); it is not explicit in the Code.

'types' for which the Code provides a list of pre-packaged terms,[23] some of which are waivable by the parties (*dispositions supplétives*) while others are not (*dispositions impératives*).[24] Even contracts initially classified as 'innominate'—contracts that fail to qualify under any of the available nominate contracts—are typically presumptively governed by the rules applicable to the nominate contract(s) that they most resemble.[25] Only once this initial stage of *classification* is completed does the court launch into the stage of *interprétation* proper, that is, turns to consider the terms of the particular contract at hand to determine the extent to which, if at all, the parties might have intended to depart from some of the *dispositions supplétives* applicable to their contract type.[26] Much of the normative content of the contract is thus already (objectively) determined by the time the court reaches the stage of *interprétation*.

To be sure, some room at that stage is made for the parties' own intentions, inasmuch as that stage begins, as suggested, with an inquiry into the extent to which the content which the parties intended to assign to their contract might differ from the objective content presumptively fixed by the Code. The intention of the parties there under examination is their *subjective*—or 'actual', 'psychological', 'private' or 'internal'—intention,[27]

[23] The section of the French Civil Code describing the 'Essential Conditions of Validity' generally applicable to all conventions is accordingly supplemented by 'the rules particular to certain contracts established under the titles relating to each one of them' (Art 1107 CC). Thereafter appears the list of 15 said titles, covering Arts 1387–2203 CC (Title IV—Marriage; Title VI—Sale; Title VII—Exchange; Title VIII—Lease; Title IX—Incorporation; Title X—Loan; Title XI—Deposit; Title XIII—Mandate; Title XIV—Suretyship; Title XVI—Settlement etc). The CCQ similarly provides for 18 nominate contracts. See Title II—Nominate Contracts of Book Five—Of Obligations, covering Arts 1708–2643 CCQ.

[24] For example, Art 1388 CC, in the Title IV—Marriage, provides that, in a marriage contract, 'the spouses cannot derogate either from the rights and obligations accruing to them as a result of the marriage or from the rules governing parental authority, legal administration, and tutorship'. The CCQ is more explicit, stating more generally that '[i]n the exercise of civil rights, derogations may be made from those rules of this Code which supplement intention, but not from those of public order' Art 9 CCQ.

[25] Nicholas, above n 1, at 49.

[26] As Nicholas aptly explains: 'The French starting-point is that the incidents of a contract are fixed by law, subject to the parties' power to vary them. … French law began with the Roman system of typical contracts and superimposed on it the unitary consensual principle that any agreement is a contract.' Nicholas, above n 1, at 49. It has been suggested that the Roman system of typical contracts was similarly reproduced in the English writ system: W Buckland and A McNair, *Roman Law and Common Law* (Cambridge, Cambridge University Press, 1936) 204ff. Although this is to a certain extent the case, the English similarity with the Roman system is much less significant for our purposes in that it, unlike the French, was more the result of a contingent combination of institutional circumstances than that of some form of conscious juridical deliberation. In addition, the establishment of the cause of action hardly determined the normative content of the parties' transaction to the same extent that it does at French law (ie, with a whole package of clearly defined terms).

[27] J Flour and J-L Aubert, *Les obligations—1. L'acte juridique*, 5th edn (Paris, Armand Colin, 1991) § 189; Terré *et al*, above n 20, at § 228; B Starck, H Roland and L Boyer, *Droit civil: les obligations*, 6th edn (Paris, Litec, 1999) 192; H Mazeaud, J Mazeaud and F Chabas,

as comparatists have often outlined.[28] That is to say, any evidence as to what the parties actually and possibly individually intended their contract to contain, explicitly or implicitly, would in principle be considered relevant, even such evidence as might not have been formally shared with the other party at the time of forming the contract. French jurists have long considered that the subjective conception of contractual intention is logically entailed by the moral ideal of the 'autonomy of the will', which traces the source of contractual obligation to the will of the individual as self-governing agent.[29] At any rate, that is the justification which they provide in support of the code dispositions that they see as embodying this ideal, in particular, the foundational article 1134 CC: 'legally formed conventions are to be considered as law by the parties who made them'.[30] As this justification for inquiring into the parties' subjective intention taps straight into the heart of the consecrating function as defined above, French contractual *interprétation* thus far can be seen as consisting of a pocket of consecrating in the otherwise unambiguously disciplining larger framework of contract law.

It has been observed that French jurists may be mistaken in thinking that the subjective conception of intention logically follows from the ideal of the autonomy of the will,[31] and that French law in practice is not nearly as

Leçons de droit civil, 8th edn (Paris, Montchrestien, 1986) 161. For Quebec authors to the same effect: J Pineau, D Burman and S Gaudet, *Droit des obligations* (Montreal: Thémis, 2006) esp §§ 1696–99.

[28] K Zweigert and H Kötz, *Introduction to Comparative Law* (Oxford, Clarendon Press, 1987) vol II, 83–94; D Harris and D Tallon, 'General Introduction' in Harris and Tallon (eds), *Contract Law Today—Anglo-French Comparisons* (Oxford, Clarendon Press, 1989) 1–5; J Cartwright, 'Defects of Consent and Security of Contract: French and English Law Compared' in P Birks and A Pretto (eds), *Themes in Comparative Law in Honour of Bernard Rudden* (Oxford, Oxford University Press, 2002) 156–7; M-A Frison-Roche, 'Remarques sur la distinction de la volonté et du consentement en droit des contrats' [1995] *Revue trimestrielle de droit civil* 573; A Rieg, *Le rôle de la volonté dans l'acte juridique en droit civil français et allemand* (Paris, LDGJ, 1961); Nicholas, above n 1, at 47–9.

[29] 'French law consecrates the omnipotence of the actual will of the author of a declaration of will. This is just the logical consequence of the principle of the autonomy of the will.' J Chabas, *De la déclaration de volonté en droit civil français* (Paris, Sirey, 1931) 81–2. See generally E Gounot, *Le principe de l'autonomie de la volonté en droit privé, étude critique de l'individualisme juridique* (Paris, A Rousseau, 1912); V Ranouil, *L'autonomie de la volonté, naissance et évolution d'un concept* (Paris, PUF, 1980); G Rouhette, 'The Obligatory Force of Contract in French Law' in Harris and Tallon, *ibid*, at 38–40. For similar remarks in Quebec doctrine: Pineau *et al*, above n 27, at §§ 35, 154–6, 223–8.

[30] '*Art. 1134. Les conventions légalement formées tiennent lieu de loi à ceux qui les ont faites* …' The equivalent CCQ provisions are Art 1378 ('A contract is an agreement of wills by which one or several persons obligate themselves to one or several other persons to perform a prestation.') and Art 1386 ('The exchange of consents is accomplished by the express or tacit manifestation of the will of a person to accept an offer to contract made to him by another person').

[31] C Valcke, 'Objectivisme et consensualisme dans le droit français de l'erreur dans les conventions' (2005) 2 *Revue de la Recherche Juridique* 661.

subjective as French jurists portray it to be.[32] If anything, though, this observation would serve to reinforce the portrayal of French contractual interpretation just given. As our objective is merely to identify the distinctively French rhetoric that emerges from the materials under study—to come to see these materials in the way that French legal actors themselves see them—the fact that their conception of contractual interpretation might somehow be mistaken or unsupported in practice is not immediately relevant. Inasmuch as the justification that French legal actors provide in support of that conception shows that *they* see it as fostering the consecrating purpose of contract law, it is possible to say that the consecrating dimension of contract law is important to the French.[33] As explained in Part I, moreover, the fact that French legal practice might sit uneasily with the subjective conception only serves to confirm the importance that they attach to that conception, for it is an example of legal actors insisting on a particular reasoning *despite* the availability of a possibly more plausible alternative (the objective conception). Precisely because the subjective conception of contractual interpretation may be questionable on logical or practical grounds, it is possible to say of French law that it '*clings* to the ideal of autonomy of the will', and that it is 'therefore 'subjective' [*only*] in terms of ideology and rhetoric'.[34]

Even such a small pocket of consecrating might be more than one should care to admit to, however. For that pocket is not just small: its very existence, as well as the modalities for establishing this existence, is ultimately conditioned by the discipline of the larger framework. Like all legal systems, French law contains rules of evidence that constrain the means by which the parties' (subjective) intention can be established in court, many of which much resemble various strands of the parol evidence rule of English law.[35] If only through the imposition of evidentiary

[32] M Tancelin, *Des obligations* (Montreal, Wilson and Lafleur, 1988) § 215.

[33] The possibility of mistake here does not undermine the idealised conception of legal actors at play. See above n 8. Even if logically mistaken, the subjective conception of contractual intention is in fact supported by French legal materials, that is, it is a mistake in logic only, not in law. As such, this mistake would be one that even an idealised legal actor could make.

[34] Vogenauer, above n 2, at 127 (emphasis added). The fact that the French might 'cling' to this ideal in the face of contradicting legal practice moreover is revealing as to the relatively lower significance which they attach, more generally, to legal practice as compared to abstract moral discourse.

[35] Much like s 4 of the English *Statute of Frauds* and the English parol evidence rule more generally, Art 1341 (1) CC provides that contracts involving sums in excess of €1500 must be proven through the production of a signed or notarised document, and that such proof may not be contradicted through testimonial evidence. Some of the many exceptions to the rule of Art 1341 (1) CC are listed in the Code itself (Arts 1341(2)–1348 CC); others were developed by *la jurisprudence* (Civ 31 May 1948 S 1949 1.127) and *la doctrine* (Terré *et al*, above n 20, at § 153; J Ghestin and G Goubeaux, *Traité de droit civil: introduction générale*, 4th edn (Paris, LGDJ, 1994) §§ 663–4). Similar rule and exceptions are found in the CC Q at Arts 2860–68.

constraints, the disciplining framework retains a hold on the consecrating pocket. More significantly, however, the consecrating justification that French jurists offer in support of their reverence for party intention is presented as a *moral* justification—as one that is *extra*-legal.[36] Of course, from the moment that this justification is articulated by these jurists in the course of legal arguments, it is no longer strictly moral, particularly given that the jurists are themselves considered a quasi-formal source of law.[37] Strikingly, however, the jurists' appeal to the moral ideal of the autonomy of the will is almost always made concurrently with an appeal to article 1134 CC,[38] which suggests that they feel justified in appealing to that ideal *only in so far as it can be shown to have been embodied into objective law*.[39] That is to say, French jurists view the parties' intention as being derivatively, not inherently, legally normative—as being legally normative only in so far as it qualifies as one of those facts to which the law has decided to extend its normativity. This is confirmed by the fact that the deciphering of party intention at French law is in principle considered a (mere) issue of fact, unlike the earlier stage of *classification* of the contract, which is considered an issue of law.[40] Therefore, our small pocket of consecrating nonetheless remains thoroughly conditioned by the larger disciplining framework.

Thus far our discussion has focused upon the dominance of the disciplining function in the French conception of contract law. But that discussion also exposes the consecrating and disciplining functions as strictly delineated from one another. For subjective contractual intention—what the parties actually intended—is a conceptually clear and distinct

[36] See, eg, Flour and Aubert, above n 27, at § 94; Terré *et al*, above n 20, at §§ 19–20; Pineau *et al*, above n 27, at § 154.

[37] Above n 11.

[38] See, eg, Flour and Aubert, above n 27, at § 95; Terré *et al*, above n 20, § 5; Pineau *et al*, above n 27, at §§ 155, 223.

[39] Flour and Aubert's following comment (*ibid*) on Art 1134 (1) CC is particularly eloquent in this respect: 'There is no need to rehash here the excessive character of this 'equalisation' of the contract with the law: in reality, the latter is superior to the former, since the former's validity always is subordinated [to it].'

[40] Ch réunies 2 February 1808 S 1808 1.480; D Lluelles and B Moore, *Droit des obligations* (Montreal, Thémis, 2006) §§ 1726ff; Tancelin, above n 32, at § 228; Terré *et al*, above n 20, at §§ 459ff. This entails that the lower courts' interpretive rulings on contractual interpretation in principle are not reviewable by higher courts, including the French supreme court in matters of private law, the *Cour de cassation*. Institutional concerns that lower courts were abusing this 'sovereign power' (*pouvoir souverain*) of theirs however led to the institution of the *recours en dénaturation*, which allows for higher court review in cases where the lower court's interpretation of 'clear and precise terms' is so outrageous as to amount to a 'distortion' (*dénaturation*) of the contract. Civ 15 April 1872 S 1872 1.232. Such *recours* remain highly exceptional, however: J Voulet, 'Le grief en dénaturation devant la Cour de cassation,' *Jurisclasseur Périodique* 1971.1.2410, [1]; M-H Maleville, *Pratique de l'interprétation des contrats: étude jurisprudentielle* (Rouen, Publications de l'Université de Rouen, 1991) §§ 149ff.

notion, certainly tighter than its objective counterpart, the somewhat loose[41] notion of what the parties might, or ought to, or could reasonably have intended. This is the case whether one considers explicit or implicit subjective intention, in spite of the obvious practical difficulties attending an investigation of implicit subjective intention. While subjective intention, especially in its implicit incarnation, indeed may at times be difficult to access, its determination remains free of the kind of substantive indeterminacy that besets an inquiry into objective intention. No question arises as to what is to count as 'subjective intention'. Such intention, being a matter of fact to be determined through empirical investigation, all and only that which was in fact intended by the parties is to count as 'subjective intention' for the purpose of legal consecration. The strict delimitation of the pocket of consecrating from the surrounding disciplining framework is hence made possible by the conceptual tightness of the notion upon which it is built, namely, party intention as real, empirically determined intention.

That the French regard the pocket of consecrating as strictly delimited as being borne out by the fact that, as indicated, the determination of party intention at French law is formally earmarked an issue of fact, whereas the other determinations making up the contractual enforcement process are earmarked issues of law. This also is evident from the rules governing the means by which party intention is to be accessed and established in court proceedings, the package of which is, as suggested, functionally roughly equivalent to the parol evidence rule of English law. The French and English packages indeed differ in their respective organisations. Unlike the English package, which as a whole straddles evidentiary and substantive law, the French package neatly divides into its substantive and evidentiary components. Rules pertaining to the classic canons of interpretation, most of which were designed to guide the interpreter in accessing party intention (the *contra proferentem* rule, the rule *ut res magis valeat quam pereat* and so on), are considered rules of substantive law and accordingly found in the section of the Code entitled 'Of the Interpretation of Conventions' located in one of the chapters devoted to substantive contract law.[42] In

[41] That one notion is described as 'tight' and the other as 'loose' should not be read as a suggestion that the one is somehow 'better' or 'superior' or 'more desirable' than the other. Conceptual looseness may be indicative of greater complexity and richness, and it may entail greater flexibility, which is generally considered a virtue from a legal perspective. Conversely, conceptual 'tightness' may be symptomatic of shallowness, bareness or rigidity. Simplistic notions are conceptually tight, yet are no less defective for that.

[42] Arts 1156–64 CC, contained in Section V of Chapter III of Title III (*De l'interprétation des conventions*). The same rules are found in Section IV ('Interpretation of Contracts') of Chapter II ('Contracts') of Book Five ('Obligations') of the CCQ. See, eg, Art 1427 ('Each clause of a contract is interpreted in light of the others so that each is given the meaning derived from the contract as a whole.'); Art 1428 ('A clause is given a meaning that gives it some effect rather than one that gives it no effect.'); Art 1430 ('A clause intended to eliminate doubt as to the application of the contract to a specific situation does not restrict the scope of

contrast, rules pertaining to evidence *per se*, form a chapter of their own entitled 'Of the Proof of Obligations, and of their Payment'.[43] French doctrinal treatments of these issues moreover exhibit the same segregating tendency: it is not uncommon to find purportedly general discussions of contractual interpretation that are completely silent on evidentiary issues,[44] and where those issues are discussed, it is typically in a section separate from that devoted to the substantive rules.[45] Conversely, no French treatise on the law of evidence that could be found even alludes to the substantive rules.[46]

The content of the substantive rules further confirms the conceptual tightness of subjective intention and resulting strict delimitation of the consecrating pocket from the surrounding disciplining framework. Unlike the evidentiary rules, which, as explained, cannot but be disciplining in nature, the substantive rules contained in the interpretation section of the Code are not intended to modify, add to, or subtract from the parties' actual intention in any way. Consistently with the empirical conception of intention as made up of all and only that which the parties actually intended, those 'rules' are in fact mere suggestions, interpretive guidelines,[47] codified for no other purpose than to assist the courts in their deciphering task.[48] In the same vein, French law has never placed a priori limitations on the admissibility of materials deemed relevant for that purpose: as alluded to above, all materials likely to help shed light on the parties' actual intention in principle are admissible, including materials pertaining to such contextual factors as pre-contractual, collateral, or

a contract otherwise expressed in general terms.'); Art 1432 ('In case of doubt, a contract is interpreted in favour of the person who contracted the obligation and against the person who stipulated it.').

[43] Arts 1315–69 CC, contained in Chapter VI of Title III (*De la preuve des obligations et de celle du paiement*). The CCQ is even more eloquent in this regard, as it segregates the articles relating to evidentiary issues (Arts 2803–74 CCQ) into a book of their own ('Book Seven—Evidence'), altogether distinct from the general book on obligations ('Book Five—Obligations').

[44] See, eg, Maleville, above n 40; Pineau *et al*, above n 27, at § 224 (who devote four lines to evidentiary issues).

[45] See, eg, Terré *et al*, above n 20; Ghestin and Goubeaux, above n 35.

[46] See, eg, C Perelman and P Foriers, *La preuve en droit* (Brussels, Bruylant, 1981); CE Dorion, *De l'admissibilité de la preuve par témoins en droit civil* (Montreal, Whiteford & Théoret, 1894).

[47] Vogenauer explains that the German codifiers, unlike the French, deliberately refrained from including those rules in their code precisely because they considered them to be 'nothing but rules of logic'. Vogenauer, above n 2, at 130–31.

[48] 'pieces of advice given to the judges, in matters of contractual interpretation, more than strict mandatory rules from which they may never depart, not even in the most compelling of circumstances' Req 18 March 1807 S 1807 1.361 (C). See generally Maleville, above n 40, at §§ 260ff.

subsequent statements or actions by the parties,[49] even such materials as may pertain to one party's intention undisclosed to the other.[50] This explains why the debate over literal versus contextual interpretation has, in France unlike in England, focused almost exclusively upon statutory, not contractual, interpretation:[51] as article 1156 CC bluntly confirms,[52] literalism simply is antithetical to the subjective conception of contractual interpretation. This also explains why French law does not have anything resembling the English doctrine of rectification: there is no need for some corrective measure to bring contractual interpretation in line with party intention in a system where these are one and the same.[53] A small pocket of consecrating contained in and conditioned by a larger disciplining framework, then, yet one that can nonetheless be clearly delineated from that framework.

The strict delineation between the consecrating and disciplining functions, and the dominance of the latter, are further reiterated in the second and final phase of the interpretation stage. The reader will remember that, following the operation of *classification* of the contract, at the stage of contract formation, the court moves to the interpretation stage proper, the first phase of which, as just explained, consists in the court deciphering the parties' explicit and implicit intention. The second phase involves the court determining what to do with that intention. In particular, the court then determines which parts of that intention ought to be retained, which ought to be discarded, and how that intention is to be supplemented, all with a view to processing that intention into something worthy of legal authority. As indicated, of the many provisions presumptively assigned to their contract type under the Code, the parties are allowed to depart only from those deemed *dispositions supplétives*, but not from those deemed *dispositions impératives*.[54] The court accordingly includes party departures from

[49] '[One must] consult at the same time the terms of the act, the circumstances that preceded it, those that followed it, …' Req 9 May 1877 DP 1878 1.30, cited in Maleville, above n 40, at § 403. See generally, §§ 403–42. The same is true under Quebec law: Lluelles and Moore, above n 40, at §§ 1599ff.

[50] Above text accompanying nn 27–9.

[51] P-A Côté, *Interprétation des lois* (Montreal, Yvon Blais, 1982) 18ff. To the extent that the teachings of the *exégètes* did transpire from statutory to contractual interpretation, they would be reflected in the exceptional remedy of the *recours en dénaturation*. See Vogenauer, above n 2, at 131–2.

[52] 'One must in contracts seek to ascertain what was the common intention of the contracting parties, rather than stop at the terms' literal meaning.' The equivalent CCQ article reads: 'The common intention of the parties rather than adherence to the literal meaning of the words shall be sought in interpreting a contract.' Art 1425 CCQ.

[53] See, eg, Ghestin's discussion of the famous German case involving parties who meant to buy and sell whale meat but wrote 'shark meat' in the contract: J Ghestin, *Traité de droit civil: la formation du contrat*, 4th edn (Paris, LGDJ, 2000) § 495.

[54] Above n 24.

the former, but not party departures from the latter. The court also disregards clauses otherwise deemed 'abusive',[55] which the Civil Code of Quebec (CCQ) defines as follows:

> An abusive clause is a clause which is excessively and unreasonably detrimental to the consumer or the adhering party and is therefore not in good faith; in particular, a clause which so departs from the fundamental obligations arising from the rules normally governing the contract that it changes the nature of the contract is an abusive clause.[56]

At French law like at Quebec law, moreover, the notion of an 'abusive' clause is semantically—thus also juridically[57]—related to the notion of 'abuse of rights', the disciplining flavour of which is such as to startle any English lawyer. The latter notion indeed operates as a license for the courts to find that particular actions by the parties, although technically within the ambit of their rights, nonetheless are to be denounced as illegitimate *exercises* of these rights.[58]

Following article 1135 CC, moreover, the court also reads into the contract 'all that ought to follow from the nature of the obligation under equity, usage, and the law'.[59] The main such obligation 'following from equity' is the famous obligation to execute contractual obligations in good faith, codified at article 1134(3) CC[60]—'a powerful filter'[61] for onerous

[55] Civ 1ère 6 déc 1989 D 1990.289 note J Ghestin, RTD civ 1990.277 obs J Mestre. The legal basis for that power is still contested: some judicial decisions invoke Art 1134 CC, while others refer to Art 35 of the *Loi du 10 janvier 1978*.

[56] Art 1437 CCQ.

[57] The semantic-juridical connection flows from the deliberateness of the codification process.

[58] See, eg, *National Bank of Canada v Soucisse* [1981] 2 SCR 339 (failure of bank to inform heirs of revocability of suretyship provided by deceased); *Houle v Canadian National Bank* [1990] 3 SCR 122, 74 DLR (4th) 577 (bank recalling loan and realising on its guarantee within hours of announcing its intention to do so). The French notion of abuse of rights somewhat resembles, yet is not nearly as strictly framed as, the estoppel of English law. See generally A Mayrand, 'Abuse of Rights in France and Quebec' (1973–74) 34 *Louisiana Law Review* 993.

[59] The equivalent CCQ article reads: 'A contract validly formed binds the parties who have entered into it not only as to what they have expressed in it but also as to what is incident to it according to its nature and in conformity with usage, equity or law.' (Art 1434 CCQ.)

[60] The equivalent CCQ article reads: 'The parties shall conduct themselves in good faith both at the time the obligation is created and at the time it is performed or extinguished.' (Art 1375 CCQ.)

[61] P Delebecque and D Mazeaud, 'Les clauses de responsabilité: clauses de non-responsabilité, clauses limitatives de réparation, clauses pénales' in M Fontaine and G Viney (eds), *Les Sanctions de l'inexécution des obligations contractuelles* (Paris, LGDJ, 2001) 361, 372.

contractual clauses. As articles 1134 and 1135 are located in the *Dispositions générales* of Chapter III titled 'Of the effect of obligations', they are generally applicable to all contracts.[62]

The disciplining dominance at play in this second phase of the interpretive stage is obvious enough, in so far as this is the moment at which the law's transformative power is applied to the parties' intention. The intention that ends up being legally consecrated is the parties' original intention as filtered, supplemented, or at the very least endorsed by the court. Of our original pocket of consecrating indeed not much is left. Many French scholars accordingly have described this second phase in such terms as '[the stage at which] the common will of the parties is seized as a pretext for justifying holding the parties to the rules of equity'.[63] Further still, in many cases the justifications offered for such judicial interventions do not even allude to the common will of the parties, whether as pretext or otherwise.[64] More often than not, those justifications explicitly appeal to such unambiguously disciplining notions as 'the court's desire to protect the weaker party'.[65] As some prominent observers have concluded, '[i]t therefore could not be better established that these [contractual] obligations draw their source, not from some professed will of the parties, but from the law'.[66]

This second phase also underscores the French insistence on keeping the consecrating and disciplining functions strictly delineated. However little may be left after the second (disciplining) phase of the parties' original intention identified in the first (consecrating) phase, French jurists insist on keeping these phases separate, not just sequentially, but also semantically:

[62] The obligation of good faith contained at Art 1375 CCQ is even more general, as it is found in the 'General Provisions' applicable to all obligations, not just contractual obligations. Moreover, it explicitly governs all the stages in the life of the obligation, not just its execution. See above n 60. Judicial and doctrinal interpretations of the French Art 1134 (3) CC have likewise extended the ambit of that disposition to cover all stages of the life of contractual obligations. See, eg, Terré *et al*, above n 20, at §§ 177, 414.

[63] Maleville, above n 40, at § 453, fn 23, citing Boyer, Gaudemet, Mazeaud and Mazeaud, Flour and Aubert, Dupichot, Dereux, Chabas, Lopez Santa Maria and Salle de la Marnière.

[64] H Beale, 'Exclusion and Limitation Clauses in Business Contracts: Transparency' in Burrows and Peel, above n 2, at 191, 200; Ghestin, above n 53, at § 407; Delebecque and Mazeaud, above n 61, at 388–90.

[65] Terré *et al*, above n 20, at § 297 (discussing the Cour de cassation *jurisprudence* of the second half of the nineteenth century, in which clauses providing for the payment of steep professional fees in contracts for professional services were almost systematically struck down); Delebecque and Mazeaud, above n 61, at 363 ('[the courts] having considered null conventions of non-liability, precisely for moral reasons and by reason of the fundamental value attached to the idea of fault, …'); Cass civ 2e 17 February 1955 D 1956.17 ('are null the clauses excluding or limiting liability in delictual matters, given that articles 1382 and 1383 of the Civil Code are of public order and their application hence not possibly paralyzed by a convention').

[66] Terré *et al*, above n 20, at § 456.

the first is designated as 'subjective',[67] 'real'[68] or 'explicative'[69] interpretation whereas the second is respectively referred to as 'ideal', 'divinatory' or 'creative' interpretation, among other pairs of contrasting labels.[70] Such insistence is consistent with the different locations in the code of the rules that govern these two phases respectively. The disciplining rules of articles 1134 and 1135 CC, being found among the *Dispositions générales* of Chapter III, are kept separate from the consecrating rules pertaining to the classic canons of interpretation contained in the *De l'interprétation des conventions* section. The strict delineation of the consecrating and disciplining functions, just like the overall dominance of the latter, is therefore reflected in both phases of the interpretive stage.

Hence the view of contract law that emerges from the French materials on contractual interpretation is that of an overall disciplining framework with a pocket of consecrating, which remains small and subject to the discipline of the larger framework. The conception of contractual intention as actual, empirical intention which grounds this pocket accounts for its being both ultimately subjected to the discipline of the larger framework and nonetheless clearly delineanable. As an empirical matter, subjective contractual intention cannot but draw its normative significance from the surrounding legal framework, which in turn sets the terms upon which such significance is to be supplied. Whereas these terms are merely evidentiary in the first, 'deciphering' phase of the interpretive process, they also are substantive in the second, 'processing' phase. But subjective intention as an empirical matter also exhibits sufficient conceptual tightness to sustain the integrity of the pocket as against the larger framework. In that sense, the view of contract law that emerges from the civil law materials on contractual interpretation can be described as that of a process in which the disciplining function is far more prominent than the consecrating one, yet the two functions remain neatly delineated from one another.

III. (ENGLISH) COMMON LAW MATERIALS ON CONTRACTUAL INTERPRETATION

In contrast to the civilian view of contractual interpretation, the view of contract law that emerges from the common law materials is that of a process in which the consecrating and disciplining functions are not clearly

[67] Maleville, above n 40, at § 452.
[68] Flour and Aubert, above n 27, at § 396.
[69] Terré *et al*, above n 20, at § 325.
[70] Pineau *et al* (above n 27, at 989) use the labels 'explicit' v 'implicit' interpretation, which is perhaps less apt as the first phase includes an investigation of explicit *and* implicit party intention.

delineated, more specifically, in which the disciplining function tends to be folded into the consecrating function in an attempt to downplay the former's relative significance. I propose to show that this is the case by raising, and ultimately dismissing, some potential counter-examples, that is, by identifying the rules and institutions of English law that most appear to embody a clear consecrating/disciplining distinction and/or a disciplining dominance, and then demonstrating that they do so in appearance only. These rules and institutions are the jurisdictional division between law and equity as it pertains to contracts; the law/fact distinction; the doctrine of implied terms; the particular treatment of onerous contractual terms; and the doctrine of frustration and certain elements of the law of remedies.

The jurisdictional division between law and equity—probably the most fundamental organising feature of English private law—at first sight seems as if it might supply a consecrating/disciplining delineation similar to that which animates French contract law. Under that division, the courts of law generally were to enforce contracts as expressed by the parties, whereas the courts of equity were to set the contract aside where justice would demand it.[71] If only from a jurisdictional perspective, therefore, the consecrating and disciplining functions seemingly were to be neatly compartmentalised.

The law/equity division never really tracked the consecrating/disciplining distinction elaborated in Part I, however. For the contractual intention that the law courts have had to enforce is the parties' *objective* intention[72]—the intention 'which the party in question by his actions or words displays to the other, not some hidden intention which he may have concealed in the inner reaches of his mind'.[73] And as suggested, objective intention is a relatively loose concept, at least in comparison with subjective intention. At the outset, the very objective/subjective distinction at English law is elusive at best. Although objective and subjective elements are clearly both at play in the English conception of contract (albeit with a clear predominance of the former), their respective domains have yet to be reasonably neatly delineated.[74] Most significantly, however, the 'objective' intention of

[71] See the chapter entitled 'Of the Equitable Jurisdiction in Relieving Against Unreasonable Contracts or Agreements' in *Powell on Contracts* (1790), cited in S Waddams, *The Law of Contracts*, 4th edn (Toronto, Canada Law Book, 1999) 319.

[72] See Blackburn J's famous statement in *Smith v Hughes* (1871) LR 6 QB 597, 607 ('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'). More recent restatements include: Lord Reid in *McCutcheon v David MacBrayne Ltd* [1964] 1 WLR 125 (HL) 127; Lord Diplock in *Ashington Piggeries Ltd v Christopher Hill Ltd* [1972] AC 441 (HL) 502; Lord Steyn in *Deutsche Genossenschaftsbank v Burnhope* [1995] 1 WLR 1580 (HL) 1587 [*Burnhope*].

[73] Abella and Rothstein JJ in *Double N Earthmovers*, above n 2, at [65].

[74] For a suggestion that contract formation pertains to the subjective domain, whereas contract interpretation relates to the objective domain, see: SA Smith, *Contract Theory*

English law itself appears to combine both kinds of elements. English courts indeed commonly appeal to the shady notion of 'the parties' *reasonable* intention', without specifying whether by that they mean 'what the parties reasonably *ought to have intended*' or 'what the parties reasonably *ought to be taken to have intended*'. The difference between these two meanings is critical, as the first betrays a clear disciplining stance, whereas the second can be seen as strictly consecrating. Yet English courts have been reluctant to clarify which of these they mean, arguably because they in fact have been meaning *both*: the parties' 'reasonable intention' stands for the intention which it is reasonable for them to have *precisely because* that is the intention which it is reasonable for each of them to attribute to the other. In other words, it is because a particular intention reasonably can (factually) be attributed to the parties that the court will endorse that intention as that which reasonably can (legally) be attributed to them. Hence does it make sense for an English court to affirm in one and the same breath that 'a release cannot *apply, or be intended to apply* to circumstances of which a party had no knowledge at the time he executed it'.[75] The same interplay of factual and normative—of consecrating and disciplining—is reflected in such statements as: '[t]he purpose [of interpretation] is to enable the Court to *attribute the appropriate objective meaning to the words* used by the parties in the document'.[76] And it is because of this interplay, I would argue, that contractual intention as objective intention is inherently unstable, and hence conceptually looser than subjective intention.

The conceptual looseness of objective intention arguably is part of the explanation for the traditional prominence of literalism and the parol

(Oxford, Oxford University Press, 2004) 174; K Lewison, *The Interpretation of Contracts* (London, Sweet & Maxwell, 2004) 28–31. Recent case law on contractual interpretation however veers dangerously close to the subjective domain. See in particular Lord Hoffmann's recent suggestion that it is only for reasons of 'practical policy' that the parties' subjective knowledge is excluded from the matrix of facts which judges ought to consider when interpreting contracts (*Investors Compensation Scheme Ltd v West Bromwich Building Society et al* [1998] 1 WLR 896 (HL) 913 [*ICS*]). See also Lord Nicholls, 'My Kingdom for a Horse: The Meaning of Words' (2005) 121 *LQR* 577, 586–9 (pleading for the admissibility of previous negotiations and declarations of intentions for the purpose of interpreting contracts); and Rix J (as he then was) in *BHP Petroleum Ltd v British Steel plc* ([1999] 2 Lloyd's Rep 583 (QBD), affirmed in part: [2000] 2 Lloyd's Rep 277 (CA) (suggesting that the *Hadley v Baxendale* test was 'conceptually difficult' in so far as it entails that where a loss normally falling under the second branch of the test is in fact known to the defendant it becomes knowable by a reasonable person in the position of the defendant, and thus recoverable under the first branch of the test).

[75] Pollock CB in *Lyall v Edwards* (1861) 158 ER 139 (QB) 143 (emphasis added).

[76] Hobhouse LJ in *Credit Lyonnais Bank Nederland NV v Export Credit Guarantee Department* [1998] 1 Lloyd's Rep 19 (CA) 38 (emphasis added). Another example is Lord Steyn's statement that '[t]he purpose of interpretation is to assign to the language of the text the most appropriate meaning which the words can legitimately bear'. *Equitable Life Assurance Society v Hyman* [2000] 1 AC 408 (HL) 458 [*Hyman*].

evidence rule in English contract law—the other part pertaining to such notorious institutional factors as the endeavour to keep contractual interpretation matters away from the (often illiterate) juries.[77] Although neither literalism nor the parol evidence rule is logically entailed by objective intention,[78] their emergence is hardly surprising in a context where unconstrained judges have historically been disapprovingly equated with legislating judges and certainty in transactions is generally highly prized. As the notion of objective intention, in its looseness, would not provide the requisite certainty and constraining power, these would be supplied through extraneous rules, for example, courts ought not look beyond 'the four corners of the document' when interpreting contracts, and words ought to be given their narrowest, most literal meaning.[79] Just as the rule of precedent arose as an effective means by which to curb the judges' legislative leanings and to restore certainty, so literalism and the parol evidence rule emerged as effective means tailored for contractual interpretation.

In turn, the need arose for a doctrine along the lines of rectification to correct the abuses generated through blind adherence to literalism and the parol evidence rule.[80] Hence, one could exceptionally bypass the four corners of the document as well as the literal meaning of the words in cases where exclusive reliance on them would be unfair since neither reflects the parties' actual intention. If in some cases hardship or injustice may be effected by this rule of law, such hardship or injustice can generally be obviated by the power in equity to reform the contract, in proper cases and on proper evidence that there has been a real intention and a real mistake in expressing that intention; these matters may be established, as they generally are, by extrinsic evidence. The court will thus reform or rewrite the clauses in order to give effect to the real intention. However, that is not construction, but rectification.[81]

In the context of contractual interpretation centred upon objective intention, therefore, the purportedly respectively consecrating and disciplining roles of law and equity are at times reversed: interpretation of the document at law plays the consecrating role where the document reflects

[77] *Pioneer Shipping Ltd v BTP Tioxide Ltd* [1982] AC 724 (HL) 736 (Lord Diplock) [*Pioneer Shipping*].

[78] In strict logic, it indeed is possible to say, as the French do, that literalism and the parol evidence rule pertain to the kind of evidence that is deemed useful for the purpose of getting to contractual intention however substantively defined.

[79] J Steyn, 'Written Contracts: To What Extent May Evidence Control Language?' (1988) 41 *Current Legal Problems* 23; LH Hoffmann, 'The Intolerable Wrestle with Words and Meanings' (1997) 114 *South African Law Journal* 656; Lewison, above n 74, at 11.

[80] See generally J Steyn, 'The Intractable Problem of the Interpretation of Legal Texts' (2003) 25 *Sydney Law Review* 5, 8; Nicholas, above n 1, at 47–8; A Burrows, 'Construction and Rectification' in Burrows and Peel, above n 2, at 77, 78.

[81] Lord Wright in *Inland Revenue Commissors v Raphael* [1935] AC 96 (HL) 143.

the parties' intention, whereas alteration of the document at equity plays that role where the document is not reflective of party intention.[82] Conversely, it is possible to describe both these operations as playing a disciplining role: where the court favours an objective interpretation of the contract that demonstrably diverges from the parties' subjective intention, it can be said that the court imposes its own interpretation upon the parties, and where the court instead looks beyond the document to reach to the parties' subjective intention, it likewise can be said that the court is then reading into the contract the words 'which it considers ought to have been used'.[83] If anything, therefore, the law/equity distinction cuts across, rather than tracks, the consecrating/disciplining distinction elaborated in Part I.

Law and equity have since formally merged, and the ripple effect is still being felt in all fields of English law. That merger contributed little in the way of disentangling the consecrating and disciplining functions, however, arguably because contractual interpretation remains anchored in objective intention. Along with the gradual disappearance of juries in private law matters and the concurrent consolidation of the rules of law and equity, strict adherence to literalism and the parol evidence rule was eventually overcome.[84] The parol evidence rule has come to be considered as more than just a rule of evidence: 'the so-called parol evidence rule … [is not] a rule of evidence (though, like all legal rules, it has evidentiary consequences): it is a rule of substantive contract law, namely, that extrinsic statements do not affect the parties' obligations'.[85] Whereas a similar fusion of the evidentiary and the substantive would, as discussed, be highly unlikely in a legal system based on subjective intention, it is to be expected in one that centres upon objective intention. For under the objective conception of intention, evidence of intention is, as between the contracting parties, in many respects as significant as intention itself: the only *relevant* intention is that of which the other party has been given evidence.

[82] '[B]oth techniques are concerned to ensure that the contract, as enforced, reflects the intentions of the parties. The construction of a contract achieves this by *interpretation* of the words used. Rectification achieves this by *alteration* of the words used.' Burrows, above n 80, at 77.

[83] R Calnan, 'Construction of Commercial Contracts: A Practitioner's Perspective' in Burrows and Peel, above n 2, at 17, 19 (criticising Lord Hoffmann in *ICS*, above n 74.)

[84] '[I]n the case of commercial contracts, the restriction on the use of background has been quietly dropped.' Lord Hoffmann in *Mannai Investment Co Ltd v Eagle Star Life Assurance Co Ltd* [1997] AC 749 (HL) 779. 'Accordingly all the reasonably available and relevant background information is now admissible for the purposes of construing private law documents.' G McMeel, 'The Principles and Policies of Contractual Construction' in Burrows and Peel, above n 2, at 27, 45. For a recent Canadian relaxation of the parol evidence rule, see *Gallen v Allstate Grain Co Ltd* (1984) 9 DLR (4th) 496 (BCCA).

[85] Waddams, above n 71, at 225–6. To the same effect: R Stevens, 'Objectivity, Mistake and the Parol Evidence Rule' in Burrows and Peel, above n 2, at 101, 107: '[P]roperly understood it is not, or at least is no longer, a rule of evidence but a rule of construction.'

In that sense, the inherent conceptual looseness of objective intention carries over to the qualification of the parol evidence rule, as it does more generally to any line-drawing exercise as between evidentiary and substantive matters. Thus the parol evidence rule, just like objective intention, far from exemplifying our consecrating/disciplining distinction, does in fact run counter to it:

> the 'parol evidence' rule is not, as it is sometimes portrayed, a rigid rule based upon a now unfashionable love of certainty which will lead to the frustration of the parties' actual intentions. Rather the effect of the rule is to enforce, not frustrate, the intentions of the parties as objectively manifested.[86]

As for the movement away from literalism, it likewise appears to have carried with it the conceptual looseness of objective intention. Having long struggled from under the literalist legacy, English courts nowadays appear to be generally agreed that contractual words ought to be given their 'natural and ordinary meaning'.[87] In contrast with the literal approach, the current 'natural and ordinary meaning' approach—also known as 'contextual', 'purposive', 'commonsensical' or 'commercial'[88]—considerably widens the 'matrix of facts' which judges ought to consider when interpreting contracts, to the point that 'absolutely anything which would have affected the way in which the language of the document would have been understood by a reasonable man' ought now be included.[89] Just like the notion of objective intention, 'natural' meaning appears to combine a variety of judicial meanings[90] among which are the *linguistically* natural meaning (the meaning of dictionaries),[91] and the *legally* natural meaning

[86] Stevens, *ibid*, at 108.

[87] *ICS*, above n 74; *Bank of Credit and Commerce International v Ali* [2002] 1 AC 251 (HL) [*BCCI*] (whether an employee releasing 'all claims' against an employer in an employment termination agreement covers claims not known to exist in law at the time of signing the agreement.)

[88] McMeel, above n 84, at 40.

[89] *ICS*, above n 74 at 913. The relation between the consolidation of law and equity and the movement towards greater interpretive flexibility is vividly illustrated by that case. Lord Hoffmann's extension of the matrix of facts to be considered in interpreting contracts went so far as to blur the line between general principles of contractual interpretation and the equitable claim of rectification. The blurring is particularly striking in the following passage, highly reminiscent of rectification yet meant as a comment on contractual interpretation generally: '[If it seems] 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 Hoffmann in *ICS*. For more detailed discussions, see Calnan, above n 83, at 19; Burrows, above n 80, at 77.

[90] J Carter and E Peden, 'The "Natural Meaning" of Contracts' (2005) 21 *Journal of Contract Law* 277. Carter and Peden list three different meanings, not just two, although it is not entirely clear how the third, the 'application to the contract' meaning, differs from the first two: McMeel, above n 84, at 31.

[91] For example, Lord Mustill in *Charter Reinsurance Co Ltd v Fagan* [1997] AC 313 (HL) 384 (' [the meaning of the words] in the sense of their primary meaning in ordinary speech'). The House of Lords there unanimously rejected the argument that the words

(the meaning of precedents).[92] Also, as with the notion of objective intention, judges seem content to remain elusive as to which of the two meanings of 'natural' they are appealing,[93] presumably for the same reason they do not view these meanings as clearly distinct. The general movement towards greater interpretive flexibility therefore did not contribute to disentangling the consecrating and disciplining functions any more than did the drift of the parol evidence rule towards substantive law. In summary, so long as contractual interpretation remains centred upon the inherently unstable notion of objective intention, the consolidation of law and equity and the attendant liberalisation of contractual interpretation is unlikely to help clarify the line between the consecrating and disciplining functions of English contract law.

Nor is the fact/law distinction of English law likely to be more helpful in this regard. Unlike at civil law, where the line between the consecrating and disciplining functions is reinforced through the different labels (subjective/ideal, explicative/creative and so on) and designations (issues of fact/law) assigned to the two phases of the interpretive stage, the task of determining the normative content of a given contract at English law is indifferently labeled 'interpretation' or 'construction',[94] and treated as a question of law regardless of its being primarily driven by consecrating or disciplining

'actually paid' in a reinsurance treaty could be taken to entail that the (then insolvent) reinsured had to demonstrably discharge his liability to the underlying insured before claiming from the reinsurer.

[92] '[T]he legal effect … of the linguistic meaning which results from the application of the rules and principles of contract law to the linguistic meaning …' Carter and Peden, above n 90, at 279–80. See, eg, Lord Wilberforce in *Prenn v Simmonds* [1971] 1 WLR 1381, 1385 ('The words used may, and often do, represent a formula which means different things to each side, yet may be accepted because that is the only way to get 'agreement' and in the hope that disputes will not arise. … [In such a case, one must] try to ascertain the 'natural' meaning.').

[93] Carter and Peden, above n 90, at 279–80. See, eg, *Pagnan SpA v Tradax Ocean Transportation SA* ([1987] 1 All ER 81 (QBD), where Steyn J found that the 'natural meaning' of the words used by the parties in a contract for the sale of tapioca pellets for importation into the then EEC was 'more consistent' with a strict obligation on the part of the sellers to provide the requisite export certificates than with merely an obligation of reasonable diligence on their part. The Court of Appeal ostensibly agreed ([1987] 3 All ER 565 (CA)), but then proceeded to apply a default rule to determine the legal (as opposed to linguistic) effect of the words. That rule impliedly applied because it derived from the applicable law and had not been explicitly waived by the parties. 'Ultimately, the conclusion was one of law based on precedent rather than 'natural' meaning.' Carter and Peden, above n 90, at 281.

[94] 'The expression "construction", as applied to a document, at all events as used by English lawyers, includes two things: first the meaning of the words; and secondly their legal effect, or the effect which is to be given to them.' Lindley LJ in *Chatenay v Brazilian Submarine Telegraph Co Ltd* [1891] 1 QB 79 (CA) 85. See also McMeel, above n 84, at 32; Carter and Peden, above n 90, at 279. Sometimes 'interpretation' is used as a subset of 'construction' to designate interpretation of the express terms of the contract, in opposition to the 'implication' of terms into an agreement: Lord Steyn in *Hyman* above n 76, at 408, 458–9 (CA and HL).

motivations.[95] At the same time, it is widely agreed that '[t]he meaning of an ordinary word of the English language is not a question of law'.[96] Inevitably, therefore, the question of law that is the interpretation of contracts nonetheless is one that cannot but be interspersed with questions of fact.[97] Doubts have even been raised as to whether the designation of contractual interpretation as a question of law is applicable beyond the realm of written contracts.[98] Here again, the ambivalence arguably reflects the interplay of factual and normative inherent in contractual intention as objective intention. In light of the normative dimension of objective intention, the interpretive process could hardly be designated as merely 'explicative' and a question of fact. At the same time, the designation as 'normative' and a question of law is somewhat misleading given the incontrovertible factual rooting of that intention, as already explained.

The discussion of Part III thus far has aimed to expose the difficulty of disentangling the consecrating and disciplining functions at English law. But the entanglement of these functions is not symmetrical. It is not the case, in other words, that these functions are interwoven into one another so as to cause the resulting fabric to be 'neutral,' namely, neither dominantly consecrating nor dominantly disciplining. Rather, as we suggested at the outset, the disciplining function tends to be folded into the consecrating function in an apparent attempt to downplay the former's relative significance. That is to say, the consecrating function is put forward as dominant, yet various strands of disciplining at times emerge from behind that front. The contrast with French law thus lies, not merely in the difficulty in disentangling the two functions, but also in the overall consecrating rather than disciplining dominance, which likewise can be traced to the grounding notion of objective intention.

As explained, objective intention combines the respectively factual and normative dimensions of the two functions. As a particular conception of contractual *intention*, however, it itself is first and foremost a consecrating notion. However much disciplining may be at play within the notion of objective intention, therefore, it remains disciplining that is nestled within a consecrating shell. Thus, whereas at French law contractual interpretation unfolds as a disciplining drama with a cameo appearance by a

[95] As per Lord Diplock in *Pioneer Shipping*, above n 77, and in *Bahamas International Trust Co Ltd v Threadgold* [1974] 1 WLR 1514 (HL).

[96] Lord Reid in *Cozens v Brutus* [1973] AC 854 (HL) 855 (there commenting on an issue of statutory interpretation).

[97] 'Although the ascertainment of the meaning of a written contract is a question of law, many steps in the process of ascertaining that meaning are classified as questions of fact.' Lewison, above n 74, at 96.

[98] See Romer LJ's suggestion that the interpretation of an oral contract is 'entirely a question of fact', *Torbett v Faulkner* [1952] 2 TLR 659 (CA) 661.

consecrating character, at English law it showcases a consecrating character struggling to suppress its disciplining streak.

Take the issue-of-law designation just discussed. The fact that contractual interpretation is so designated despite objective intention also containing a factual dimension could be taken as a statement that contract law ought to be viewed as *primarily* disciplining in purpose. As this designation recently was emphatically reiterated outside the context of the jury system,[99] it indeed could reasonably be viewed as a somewhat deliberate affirmation of the courts' self-perception as a disciplining body with respect to matters of contractual interpretation. As one English judge declared, in an unmistakably disciplining tone, '[when interpreting a contract] the question to be answered always is, 'What is the meaning of what the parties have said?' not 'What did the parties mean to say?' … it being a presumption *juris et de jure* … that the parties intended to say that which they have said.'[100]

A closer look at the 'natural meaning' approach which judges adopt for the purpose of answering that question however, reveals that the nod towards the disciplining function in fact is only half hearted. Lord Hoffmann himself views the 'natural meaning' as resulting from a merger of the legal *into* the linguistic, that is, from a process whereby the legal is *made to conform to* the linguistic, not the other way around. In his dissent in *Bank of Credit and Commerce International v Ali*, he indeed exclaimed:

> [i]f 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?[101]

Further in the same case:

> [T]he general trend in matters of construction … 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 … 'Almost all the old intellectual baggage of 'legal' interpretation has been discarded.[102]

[99] *Pioneer Shipping*, above n 77.

[100] Lord Simon in *Schuler (L) AG v Wickman Machine Tool Sales Ltd* [1974] AC 235 (HL) [*Schuler*], reading from *Norton on Deeds* (1906), 43.

[101] *BCCI* above n 87, at [51]. 'How can the question of what a reasonable man in 1990 would have thought BCCI and Mr Naeem meant by using the language of an ACAS form be answered by examining what Lord Keeper Henley said in 1758 (*Salkeld v Vernon* (1758) 1 Eden 64, 28 ER 608)?' *Ibid*. Lord Bingham likewise said (at [9]) of the 'natural meaning' that it 'seems to me to be both good law and good sense: *it is no part of the court's function to frustrate the intentions of contracting parties* once those have been objectively ascertained' (emphasis added).

[102] *Ibid*, at [62], reiterating his comments in *ICS*, above n 74, at 912.

But the court's ruling in that case is even more interesting than its statements. The Court of Appeal had adopted a narrow reading of the contract and ruled against the employee on the question of construction, but had then ruled in his favour on the second question, whether the Bank on equitable grounds should be entitled to rely on the contract.[103] The House of Lords in contrast adopted a broader, 'natural meaning' reading of the contract, which led it to dismiss the Bank's appeal without having to consider the equity question. The 'natural meaning' approach thus provided sufficient flexibility for the House of Lords to do, under the heading of legal construction, what the Court of Appeal had considered it could only do under the heading of equity. The 'natural meaning' approach thus allowed the House of Lords to do all the disciplining it felt it had to do yet package its decision as a primarily consecrating operation.

To be sure, the 'natural meaning' of a contract is to be determined partly in light of the 'nature and object of the contract',[104] which in turn in principle is for the court largely to determine objectively, as a matter of law.[105] Just as the *classification* of the contract in accordance with its *objet* is considered an issue for objective determination at French law, therefore, its 'characterisation' at English law is determined, if not exclusively objectively, at least 'not [as a] pure question of construction'.[106] Thus is the 'natural' meaning of a commercial contract the 'commercially appropriate' meaning,[107] to the point where 'if detailed semantic and syntactical analysis of words in a commercial contract is going to lead to a conclusion that flouts business commonsense, it must be made to yield to business commonsense'.[108]

Closer inspection of the reasons offered by English courts in such cases however reveals that these courts, more than their French counterparts, are reluctant to see themselves as speaking 'for the law' as opposed to 'for the parties'. Unlike at French law, the very 'object' of the contract at English law often is presented in terms suggestive of what *the parties* considered

[103] [2000] 3 All ER 51, (HL) 69.

[104] *Darlington Futures Ltd v Delco Australia Pty Ltd* (1986) 161 CLR 500, (HCA) 510 ('[from] the context in which the clause appears including the nature and object of the contract').

[105] 'The commercial or business object of a provision, *objectively ascertained*, may be highly relevant: … But the court must not try to divine the purpose of the contract by speculating about the real intention of the parties.' Lord Steyn in *Burnhope*, above n 72, at 1587 (emphasis in original).

[106] McMeel, above n 84, at 36. For a clear articulation of the difference between contractual 'construction' and 'characterisation' at English law, see: *Street v Mountford* [1985] AC 809 (HL).

[107] *Prenn v Simmonds*, above n 92 ('profit' given the most commercially sensible meaning).

[108] *Antaios Compania Naviera SA v Salen Rederierna AB (The Antaios)* [1985] AC 191 (HL) 201 (time charterparty; shipowners allowed to withdraw vessel 'on any breach of this charterparty'; interpreted as 'on any *repudiatory* breach').

this object to be. In a recent case concerning an insurance policy containing a clause excluding liability for negligence, for example, Lord Hoffmann stated, in support of the clause's enforceability, that

> [n]egligence is a risk which the parties could reasonably have been expected to allocate to one party or the other, so as best to achieve the commercial objectives of the contract.[109]

The Supreme Court of Canada likewise declared, in a recent case involving a municipal contract issued by tender, that

> [t]he rationale for the tendering process, as can be seen from [the tender] documents, is to replace negotiation with competition. … But M.J.B. Enterprises makes clear that the tender documents control the contractual obligations of the parties to a tender, and Iacobucci J's observations were based on the particular documents in that case.[110]

The issue of the proper determination of the object of the contract moreover has arisen in other, equally significant contexts, for example in cases of mistake of identity. In the famous case of *Ingram v Little*, all judges agreed that the issue of whether the buyer's identity was an essential element of a sale agreement ought to be determined by reference to the object of the agreement, but disagreed as to how one determines that object.[111] The majority took the view that the buyer's identity was essential to the agreement merely because the seller had at the outset made that clear to the buyer.[112] Lord Devlin dissented, stating that the object of the agreement was to be derived from its nature as impartially determined by the court, not from what the parties in fact intended it to be.[113] On that basis, Lord Devlin would have concluded that the buyer's identity, while presumably possibly essential to a loan agreement for example, could not be essential to the kind of agreement there at issue, namely, a sale agreement. Lord Devlin's view arguably[114] was recently confirmed by the House of Lords in *Shogun Finance Ltd v Hudson*, where Lord Hobhouse determined that, the agreement at bar being a consumer credit agreement, the identity of the rogue there had to be considered an essential element regardless of the parties' own views on the matter.[115] Interestingly, this

[109] Lord Hoffmann in *HIH Casualty & General Insurance Ltd v Chase Manhattan Bank* [2003] 2 Lloyd's Rep 61 (HL) [67] [*HIH Casuality*].

[110] Abella and Rothstein JJ in *Double N Earthmovers*, above n 2, at [57], quoting from *MJB Enterprises Ltd v Defence Construction (1951) Ltd* [1999] 1 SCR 619, 170 DLR (4th) 577 at [41] (emphasis added; citations omitted).

[111] [1961] 1 QB 31.

[112] Sellers LJ, *ibid*, at 51.

[113] *Ibid*, at 68.

[114] Stevens, above n 85, at 112.

[115] [2004] 1 AC 919 (HL). See in particular Lord Walker at [185].

determination caused quite a stir in the English legal community,[116] which is itself revealing, I would suggest, as to the hold that the consecrating function retains over the English legal psyche.[117] In sum, the consecrating dominance of the 'natural meaning' approach is palpable even in its appeal to the object of the contract.

But what of contract terms that have traditionally eluded the application of the general principles of interpretation just described? English courts have at times acknowledged that clauses considered particularly onerous call for greater disciplining. Hence the tighter interpretive frameworks put in place to govern, in particular, penalties, forfeitures, clauses in restraint of trade, exclusionary clauses and releases. Some such clauses—forfeitures,[118] penalties,[119] clauses in restraint of trade[120]—have been considered so onerous as to warrant strict 'rules of law' against their enforcement, rules whereby such clauses would be unenforceable regardless of the parties' implicit or explicit intention. Other clauses—exclusionary clauses[121] and releases[122]—have, not without some hesitation,[123] been determined as less onerous yet sufficiently exacting to warrant the establishment of 'rules of construction', whereby enforcement is disallowed unless otherwise explicitly provided by the parties.

However, whereas the reasons offered in support of the establishment of rules of law admittedly are often framed in clear disciplining terms,[124]

[116] See, eg, A Phang, P-W Lee and P Koh, 'Mistaken Identity in the House of Lords' [2004] *CLJ* 24; C MacMillan, 'Mistake as to Identity Clarified?' (2004) 120 *LQR* 369; C Hare, 'Identity Mistakes: A Missed Opportunity?' (2004) 67 *MLR* 993; K Scott, 'Mistaken Identity, Contract Formation and Cutting the Gordian Knot' [2004] *Lloyd's Maritime and Commercial Law Quarterly* 292; G McMeel, 'Interpretation and Mistake in Contract Law: The Fox Knows Many Things …' [2006] *Lloyd's Maritime and Commercial Law Quarterly* 49.

[117] It admittedly is difficult to tell whether the outcry was primarily directed at Lord Hobhouse's suggestion that the contractual object is to be determined objectively or at the fact that he used the parol evidence rule to make that determination, as this distinction apparently eludes many commentators. Stevens, above n 85, at 107–10.

[118] See, eg, *Vernon v Bethell* (1762) 2 Eden 110, 28 ER 838 (forfeiture clause in mortgage contract).

[119] See, eg, *Protector Endowment Loan and Annuity Co v Grice* (1880) 5 QBD 592 (CA). [*Grice*].

[120] See, eg, *A Schroeder Music Publishing Co Ltd v Macaulay* [1974] 1 WLR 1308 (HL) [*Schroeder v Macaulay*].

[121] *Photo Production Ltd v Securicor Transport Ltd* [1980] AC 827 [*Photo Production*].

[122] *BCCI*, above n 87. In *BCCI*, the court explicitly applied the law of exclusionary clauses.

[123] *George Mitchell Ltd v Finney Lock Seeds Ltd* [1983] QB 284 (CA) [*George Mitchell*], affirmed [1983] 2 AC 803 (HL); *BCCI*, above n 87.

[124] 'The law itself will control that express agreement of the party; and by the same reason equity will let a man loose from his agreement, and will against his agreement admit him to redeem a mortgage.' *Howard v Harris* (1683) 1 Vern 190, 23 ER 406, 407; '[N]ecessitous men are not, truly speaking, free men, but, to answer a present exigency, will submit to any terms that the crafty may impose upon them.' *Vernon v Bethell*, above n 118, at 839. Similarly with respect to penalty clauses: '[E]quity in truth refused to allow to be enforced

those offered to support rules of construction are more ambivalent. The very fact that the courts ultimately settled for the rule-of-construction designation with respect to exclusionary clauses arguably confirms the consecrating preference of English courts, given that ample fairness (disciplining) grounds were available to justify the alternative, rule-of-law designation.[125] One possible (disciplining) reason for the courts nonetheless settling upon the rule-of-construction designation admittedly is their being satisfied that, absent any formal finding of unconscionability, the unfairness of the situation is either neutralised or trumped where the parties have expressly contracted around it. It is arguably more likely, however, that the courts favoured that designation because rules of construction involve only minimal judicial intervention with the contracting process in so far as these rules only require of the parties that they be explicit as to their intention. This is borne out by such ubiquitous judicial language as '[t]he Court has no right, in my opinion, to make a different contract for the parties'.[126]

That the courts' preference for the rule of law designation was motivated more by a concern to minimise judicial intervention than by any real sense that such designation better meets the demands of fairness is also borne out by recent declarations to the effect that the degree of judicial intervention entailed by rules of construction, however minimal, may still be considered excessive. Endorsing Lord Denning's comments in *George Mitchell Ltd v Finney Lock Seeds Ltd*,[127] Lord Hoffmann in *BCCI* indeed advocated that such rules be altogether abandoned, and the 'general principles of interpretation' extended even to contract terms as onerous as releases, on the ground that courts use rules of construction as cloaks behind which to hide when effectively rewriting contracts which they find to be unfair.[128] Whether the switch to the 'natural meaning' approach in all cases will in

what was considered to be an unconscientious bargain.' *Grice*, above n 119, at 596. And with respect to clauses in restraint of trade: 'if one looks at what [judges] said in the light of what they did, one finds that they struck down a bargain if they thought it was unconscionable as between the parties to it, and upheld it if they thought it was not. So I would hold that the question to be answered … is: 'Was the bargain fair?' Lord Diplock in *Schroeder v Macaulay*, above n 120, at 1315. Admittedly, institutional reasons also are at play here: Waddams, above n 71, at [449].

[125] '[E]xemption clauses … often amounted to taking with one hand what had been given with the other … [a] contracting party undertook various obligations and then provided that he was not to be liable if he failed to perform them.' Lord Hoffmann in *BCCI*, above n 87, at [66].

[126] 'allocating the risk of misfeasance or non-feasance … in a manner different from the allocation of that risk that the clause . . . envisages.' Lord Scott, dissenting, in *HIH Casualty*, above n 109, at 126.

[127] *George Mitchell* above n 123, at 296–7.

[128] 'When judges say that "in the absence of clear words" they would be unwilling to construe a document to mean something, they generally mean (as they did in the case of exemption clauses) that the effect of the document is unfair.' Lord Hoffmann in *BCCI*, above n 87, at [61].

fact result in greater transparency is far from clear in light of the contortions which the courts have, as suggested, long inflicted upon that expression. More probably, courts will continue to cling to reasons couched in consecrating language,[129] despite the fact that these all are cases where '[n]o amount of scrutiny of the words will provide an answer that can be said to be based solely on the expressed intention of the parties'.[130] Nonetheless, the move from rule of construction to the general 'natural meaning' approach, just like that from rule of law to rule of construction, itself is significant as it confirms that the courts tend to feel more comfortable speaking purportedly for the parties than openly for the law.

Such packaging of disciplining as consecrating moreover appears to extend to onerous clauses beyond just exclusionary clauses and releases. Where French judges, as indicated, regularly strike all sorts of contract terms on no other ground than that these terms qualify as 'abusive' at law, English judges typically prefer to interpret onerous terms away, to find that these terms are not *really* a part of the parties' agreement, as '[t]he more unreasonable the result the more unlikely it is that *the parties can have intended it*'.[131] These are just particular instances of a general pattern that, in the opinion of several observers, sees unconscionability (that is, disciplining) concerns underlying many more contract decisions than the deciders care to acknowledge.[132] This pattern perhaps is nowhere more obvious than in Lord Denning's notorious, repeated and failed attempts to reverse it.[133]

[129] See, eg, the suggestion that the *Canada Steamship* rules of interpretation of clauses excluding or limiting liability for negligence (cited in E Peel, 'Wither *Contra Proferentem*?' in Burrows and Peel, above n 2, at 53, 60) are 'merely guidelines to assist the court in *ascertaining the true intentions of the parties*.' Peel, *ibid*, at 61, citing *Lamport and Holt Lines v Coubro and Scrutton (M & I) Ltd (The Raphael)* [1982] 2 Lloyd's Rep 42 (CA). Elsewhere these rules have similarly been described as resting on the premise that it is 'inherently improbable that *one party … should intend* to absolve the other party from the consequences of the latter's own negligence.' *Gillespie Bros & Co Ltd v Roy Bowles Transport Ltd* [1973] QB 400 (CA) at 419 (Buckley LJ). See also: Lord Wilberforce in *Ailsa Craig Fishing Co Ltd v Malvern Fishing Co Ltd* [1983] 1 WLR 964 (HL) 966 ('The relevant words must be given, if possible, their natural, plain meaning.')

[130] Carter and Peden, above n 90, at 282–3 (emphasis in original).

[131] Lord Reid in *Schuler*, above n 100, at 251 (emphasis added), cited in McMeel, above n 84, at 43. Strikingly, McMeel describes this statement as 'a candidate for the most important principle of construction.' *Ibid*.

[132] 'Several generations of common lawyers have been educated in the belief that the common law of contracts admits no relief from contractual obligations on grounds of unfairness, or inequality of exchange. … It will be suggested, however, that the law of contracts, when examined for what the judges do, as well as for what they say, shows that relief from contractual obligations is in fact widely and frequently given on the grounds of unfairness, and that general recognition of this ground of relief is an essential step in the development of the law.' Waddams, above n 71, at 319. See more generally, Smith's discussion of the realist interpretation of consideration: Smith, above n 74, at 228ff.

[133] *Solle v Butcher* [1950] 1 KB 671 (CA) (with respect to mistake); *Lloyd's Bank v Bundy* [1975] 1 QB 326 (CA) (with respect to unconscionability generally); *George Mitchell*,

The merger of the disciplining and consecrating functions and dominance of the latter is also reflected in the judicial treatment of implicit contract terms. The implicit content of the contract at English law divides into terms implied in fact and terms implied by law,[134] the latter apparently representing 'the minimum decencies … *which a court will insist upon* as essential to an enforceable bargain of a given type'.[135] Here again it accordingly may look as if English law mimics French law in offering a clear consecrating/disciplining distinction, one in which the disciplining function is unswerving and fully acknowledged. But that once again is more apparent than real. To begin with, the place of implied terms in the overall law of contract is not the same at English law and at French law: '[t]he French starting-point is that the incidents of a contract are fixed by law, subject to the parties' power to vary them … [whereas] the starting point for a Common law court is the implied term'.[136] The very fact that English lawyers do through implied terms (the law of the parties) what French jurists do through rules (the law of the State) is revealing as to these legal actors' respective conceptions of contract law as primarily consecrating and disciplining.[137] In addition, the distinction between terms implied in fact and terms implied by law is bound to be murkier at English law, in so far as the English notion of a term implied by law, just like that of objective intention, embodies within it a combination of fact and law. The following passage from Lord Hoffmann's extra-judicial writings on implied terms indeed is highly reminiscent of our remarks concerning objective intention:

> In fact, of course, the implication of a term into a contract is an exercise in interpretation like any other. It may seem odd to speak of interpretation when, by definition, the term has not been expressed in words, but the only difference

above n 123 (with respect to exclusionary clauses); *Stockloser v Johnson* [1954] 1 QB 476 (CA) (with respect to forfeiture clauses); *British Movietonews Ltd v London & District Cinemas Ltd* [1951] 1 KB 190 (CA) 201 (with respect to frustration-like events).

[134] 'General default rules' and 'ad hoc gap fillers' according to Lord Steyn in *Hyman*, above n 76, at 458–9. Examples of term implied by law would be the implied terms of the implied duty of trust and confidence in the employment relationship in *Mahmud v Bank of Credit and Commerce International SA* [1998] AC 20 (HL) and *Johnson v Unisys Ltd* [2003] 1 AC 518 (HL) [35] (Lord Hoffmann). On implied terms generally, see: G Treitel, *The Law of Contract*, 11th edn (London, Sweet & Maxwell, 2003) 201–13.

[135] K Llewellyn, Book Review of O Prausnitz, *The Standardization of Commercial Contracts in English and Continental Law* (1938–39) 52 *Harvard Law Review* 700, 703 (emphasis added).

[136] Nicholas, above n 1, at 49. Nicholas then pursues (at 49–50): 'Where French law began with the Roman system of typical contracts and superimposed on it the unitary consensual principle that any agreement is a contract, English law reversed the process. In so far as it thinks in terms of typical contracts, it has derived them from the general principle of contract through the device of the implied term.'

[137] '[English legal actors] cling to the legal fiction that the courts do not make contracts for the parties. All must be traced back to the implicit will of the parties' David and Pugsley, above n 1, at 264.

> is that when we imply a term, we are engaged in interpreting the meaning of the contract as a whole. For this purpose, we apply the ordinary rule of contractual interpretation by which the parties are depersonalised and assumed to be reasonable.[138]

When describing the two kinds of implied terms at English law, it accordingly is more appropriate to refer to a gradual 'process of the hardening of fact into law'[139] or 'shades on a continuous spectrum',[140] than to the kind of strict qualitative distinction at play at French law. For the same reason, moreover, the realm of terms implied in fact is necessarily comparatively larger at English law: it includes all terms qualifying as proxies for the parties' *reasonable* intention, unlike at French law, where it includes only the terms that qualify as proxies for the parties' *actual* intention. English courts indeed have correspondingly tended to interpret 'terms implied at law' narrowly, as 'terms implied at *statutory* law', unlike their French counterparts, for whom 'law' there means the law of general application (the Code) as well as the law of exception (statutes).[141] But English courts have constricted the realm of terms implied by law even further by confining their application to cases of strict 'necessity', whereas terms implied are in fact instead subjected to the looser test of 'business efficacy'.[142] Finally, the justifications produced in support of terms implied by law often are embedded in strongly consecrational language. Lord Diplock himself once said of these terms that they flow from a 'presumption … that the parties by entering into the contract intended to accept the implied obligations'.[143] Just like terms implied in fact, therefore, terms implied by law apparently ultimately draw their legitimacy from having been (implicitly) intended by the parties. Despite its apparent similarity with its French counterpart, therefore, the English law distinction between terms implied by law and terms implied in fact turns out to be much less clear, and its disciplining dimension much more equivocal.

Other instances of disciplining packaged as consecrating can be found beyond the realm of 'interpretation' proper, finally. The law of frustration

[138] Hoffman, above n 79, at 662.

[139] Nicholas, above n 1, at 50. Nicholas gives the examples of implied terms relating to the sale of goods and to partnership, which began as reasonable implication, later became presumptions *juris de jure*, and were eventually codified.

[140] Lord Wilberforce in *Liverpool City Council v Irwin* [1977] AC 239 (HL) 254.

[141] See, eg, the Supreme Court of Canada's classification of the implied terms deemed included in tendering contracts as terms implied in fact (flowing from 'the presumed intentions of the parties') rather than as terms implied by law ('the legal incidents of a particular class or kind of contract'), despite the existence of case law establishing that such terms were to be implied in tendering contracts. *Double N Earthmovers*, above n 2, at [30].

[142] E Peden, 'Policy Concerns Behind Implication of Terms in Law' (2001) 111 *LQR* 459. The stricter test of 'necessity' admittedly recently has been considerably relaxed: McMeel, above n 84, at 32.

[143] *Photo Production*, above n 121, at 850. See also Lord Diplock referring to Lord Wilberforce at 851.

provides one example; the law of remedies another. Neither of these legal domains is alluded to in the present survey of French materials for the simple reason that it would never come to the mind of a French jurist to qualify either of these as bearing in any way upon contractual interpretation. The doctrine of *imprévision*—the French counterpart of the doctrine of frustration— is generally viewed as an issue that, just like that of contract formation, pertains to the objective delimitation of the contractual realm and accordingly cannot possibly be determined by the parties' own views on the matter.[144] Likewise with the law of remedies, it would never occur to a French jurist to think of the sanctions attached to the violation of contractual obligations as issues of interpretation: when such obligations are violated, it is for the law to intervene and apply whatever sanction is deemed objectively warranted.[145] By the Cartesian logic so dear to the French, the parties cannot possibly be both the sanctioned and the sanctioners. The justification most commonly offered in support of the rule limiting damages recoverable for breach to those flowing 'immediately and directly' from the breach,[146] indeed, appeals to objective notions of causation, not subjective notions of party intention.[147] The rules sanctioning violations of contractual obligations, like the doctrine of *imprévision*, therefore, at French law are viewed as clearly falling on the disciplining side of the consecrating/disciplining divide.

Not so at English law. To be sure, frustration and the law of remedies are, just like their French counterparts, typically classified under rubrics other than contractual interpretation. Moreover, one does find judicial[148] and academic[149] accounts of these as disciplining matters, which accounts very much resemble the French descriptions of *imprévision* and the *sanctions des obligations*. Here end the similarities, however. With respect to frustration, English courts have long deployed creative interpretation

[144] See generally Terré *et al*, above n 20, at §§ 445ff.

[145] As two prominent French jurists remarked with respect to exclusionary clauses: '*the contractual freedom to determine the consequences of a failure to live up to one's contractual obligations ends where the essence of the contract begins*.' Delebecque and Mazeaud, above n 61, at 378–9 (emphasis in original).

[146] Art 1151 CC; Art 1613 CCQ.

[147] See, eg, Terré *et al*, above n 20, at §§ 567ff; Pineau *et al*, above n 27, at § 464.

[148] See, eg, Cotton LJ in *Hydraulic Co Ltd v McHaffie* (1878) 4 QBD 670 (CA) 677: 'It cannot be said that damages are granted because it is part of the contract that they shall be paid: it is the law which imposes or implies the term that upon breach of a contract damages must be paid.'

[149] See, eg, Stevens, above n 85, at 105–107: 'This process does not depend upon the court's ability to imply *positively* as a matter of fact a condition precedent ('if the music hall does not exist there is no contract') or a condition subsequent ('if the music hall burns down the agreement comes to an end') into the agreement. In many cases such implication will be wholly artificial as it cannot be determined what the parties would have agreed if they had thought about the matter in advance. Rather the court's task is the *negative* one of construing the limits of the actual bargain the parties have entered into ….'

(consecrating) arguments in order to shirk the issue altogether.[150] And in cases where the doctrine has seemed inescapable, they have openly resisted applying it, while at times proving overly generous in their interpretation of *force majeure* clauses.[151] It has been suggested that the reason for such disparate treatment is that a finding of frustration causes 'the contract [to be] immediately and automatically brought to an end, *irrespective of the wishes of the parties*',[152] whereas enforcing *force majeure* clauses in contrast involves 'no danger of becoming involved in making a new contract for the parties or imposing an outcome irrespective of their wishes'.[153] Likewise with remedies, the rationale for the remoteness rule of damages, whether in its first inception in *Hadley v Baxendale*[154] or in its multiple subsequent iterations, is very much consecrational in tone, at least when compared to the causation rationale favoured by the French. This tone is particularly strong in the so-called 'second contract theory' cases, of which *Horne v The Midland Railway Co* is a prime example.[155] Blackburn J there famously affirmed that 'in order that the notice [of the plaintiff's

[150] See, eg, the House of Lords' judgment in *British Movietonews Ltd v London and District Cinemas Ltd* [1952] AC 166 (HL). The fact that the implied contract theory of frustration was altogether abandoned in the second half of the twentieth century confirms that such a consecrating twist on an unambiguously disciplining issue was too artificial to be sustainable. I am grateful to Stephen Waddams for this point.

[151] See, eg, *J Lauritzen AS v Wijsmuller BV (The Super Servant Two)* [1990] 1 Lloyd's Rep 1 (CA). (carriage of goods—contract provides that carrier will use one of two barges—*force majeure* clause allowing for cancellation in the event of 'perils or danger and accidents of the sea'—carrier performance rendered impossible as one barge sank and other was allocated to performance of another contract—carrier's frustration argument denied as their inability to perform primarily due to decision to allocate second barge to other contract—carrier's *force majeure* argument allowed (despite same reasoning applicable: impossibility to perform due to decision concerning second barge rather than 'perils of the sea.'))

[152] E McKendrick, '*Force Majeure* Clauses: The Gap between Doctrine and Practice' in Burrows and Peel, above n 2, at 233, 239, paraphrasing the Privy Council in *Hirji Mulji v Cheong Yue Steamship Co Ltd* [1926] AC 497 (PC) 505, 509 (emphasis added).

[153] *Ibid*.

[154] 'Where two parties have made a contract which one of them has broken, the damages which the other party ought to receive … should be such as … may reasonably be supposed to have been *in the contemplation of both parties, at the time they made the contract*. Now, if the special circumstances under which the contract was actually made were communicated by the plaintiffs to the defendants, and thus known to both parties, the damages resulting from the breach of such a contract … would be the amount of injury which would ordinarily follow from a breach of contract under these special circumstances *so known and communicated*…. *For, had the special circumstances been known, the parties might have specially provided for the breach of contract by special terms as the damages in that case*; and of this advantage it would be very unjust to deprive them.' *Hadley v Baxendale* (1854), 9 Exch 341, 156 ER 145, 151 (Alderson B) (emphasis added). Ironically, the court there heavily relied upon the writings of the famous French jurist Pothier! This is not the only time that an English court clumsily undertook to graft a French doctrine upon the English law. For a similar initiative with respect to contractual mistake, see: R David, 'La doctrine de l'erreur dans Pothier et son interprétation par la common law d'Angleterre' in P Matter (ed), *Études de droit civil à la mémoire de Henri Capitant* (Paris, Dalloz, 1939) 145.

[155] (1873) LR 8 CP 131 (Exch Ch) 135 (emphasis added).

particular circumstances] may have any effect, it must be given under such circumstances, as that *an actual contract arises* on the part of the defendant to bear the exceptional loss'.[156]

It is worth emphasising that the difference between French and English law here again is rhetorical, not functional. Functionally, the two systems differ little, in so far as they both allow for the possibility of contracting around the rules: *force majeure* and liquidated damages clauses indeed are permitted and commonly used in both systems.[157] With respect to both systems, therefore, it can plausibly be asserted that, in the absence of such clauses, the parties ought to be taken to have implicitly endorsed the rules of objective law—the doctrines of frustration/*imprévision*, the rules of remoteness/*dommages direct et immédiat*. From a rhetorical rather than a functional perspective, however, what is interesting is that the ubiquity of *force majeure* and liquidated damages clauses in practice has much affected the perception of the corresponding rules of objective law at English law, but not at French law. English lawyers have, in light of this practice, tended to see the rules of objective law as mere 'default rules', which can be waived at will,[158] and which accordingly bear at least some connection to interpretation issues. French jurists in contrast have quarantined such departures to the realm of (consecrating) 'exceptions' so as to prevent them from contaminating the (disciplining) 'principles' of objective law. The English tendency to blur the disciplining/consecrating line and recast the latter as the former, and the reverse French tendency to emphasise the disciplining/consecrating line and acknowledge the former's dominance, are thus palpable even beyond the realm of contractual 'interpretation' proper.

The view of contract law that emerges from the English materials on contractual interpretation hence is that of a process in which the disciplining function tends to be folded into the consecrating function in an apparent attempt to downplay its relative significance. The difficulty in disentangling the two functions originates from the fact that objective intention—the cornerstone of the English contract law—itself combines consecrating and disciplining elements, and the resulting conceptual looseness carries over to the various legal instruments pertaining directly (and

[156] *Ibid*, at 141. See also *Transfield Shipping Inc v Mercator Shipping Inc* [2008] UKHL 48, discussed in A Robertson, 'The Basis of the Remoteness Rule in Contract' (2008) 28 *Legal Studies* 172.

[157] With respect to liquidated damages, see Arts 1152–3 CC and Art 1622 CCQ. With respect to *force majeure* clauses, see: Mazeaud *et al*, above n 27, at § 581; Art 1470 CCQ.

[158] This may explain why the 'default rule' explanation of objective law propounded by law and economics scholars has been better received in common law than in civil law jurisdictions. The easier reception of law and economics scholarship in common law jurisdictions arguably is another element that confirms the dominance of the consecrating function in the common law psyche, as that scholarship generally aims to justify legal rules by appeal to individual preferences.

less directly) to contractual interpretation. The historical division between law and equity, the subsequent convergence of law and equity and attendant liberalisation in the interpretation of written contracts, the law/fact distinction, the doctrine of implied terms, the particular treatment of onerous terms, and even such institutions as the doctrine of frustration and certain elements of the law of remedies indeed show legal actors merging the two functions. This merger is far from symmetrical, however, in that English judges clearly prefer to see themselves as consecrating (speaking for the parties) rather than disciplining (speaking for the law) agents. The consecrating dominance of the grounding notion of objective intention carries over to the judicial treatment of contractual object and onerous clauses, as well as to the doctrine of frustration and the remoteness rule of damages. In light of the difficulty in disentangling the two functions, the pocket metaphor used for the purpose of describing the relation of subordination of the consecrating to the disciplining at French law hardly is apposite for describing the reverse relation at English law. The better metaphor would be 'a front of consecrating dominance from which strands of disciplining emerge from time to time'.

IV. CONCLUSION

As an exercise in comparative legal rhetoric, our analysis has focused upon the justifications, that is, the reasons offered in support of the particular choices of legal instruments deployed for the purpose of interpreting contracts at French and English law. For those reasons embody the core values that animate each system, in particular, their respective conceptions of contract law. The conception of contract law that emerges from the French materials on contractual interpretation is, largely owing to the conceptual tightness of the grounding notion of subjective intention, that of a process in which the disciplining function is by far the more prominent yet the two functions remain neatly delineated from one another. In contrast, the conception of contract law that emerges from the English materials is, owing to the conceptual looseness of objective intention, that of a process wherein the two functions are interwoven and the significance of the disciplining function is comparatively downplayed. Although these different conceptions may well be more or less 'right' or 'good' or 'desirable' from some extra-legal perspective, be it economic, philosophical or otherwise, all that matters for now is that they be *internally* accurate: that they correctly reflect the distinctive *self*-understandings that animate French and English contract law respectively.

As the reconstruction of these different conceptions simultaneously served to interconnect the materials on contractual interpretation in each system and to contrast those of the other system, I hope that this article

contributed to a deeper understanding of the rules and institutions at issue, that is, an understanding of these rules and institutions as they relate to each other *and* to those in the other system. This is what is meant above by a 'meaningful comparative understanding of law'.

5

Consideration and the Morality of Promising

ANDREW S GOLD*

I. INTRODUCTION

WHEN IS IT problematic for contract law to diverge from promissory morality? Seana Shiffrin has made a significant contribution to contract theory debates by focusing attention on this question. In a recent article, she supports an intermediate, 'accommodationist' approach to the relationship between law and morality, an approach which suggests that the law should be responsive to moral concerns without actively enforcing interpersonal morality as such.[1] According to Shiffrin, moral agents should not be placed in a position where they have to accept a conflict between the justifications for legal doctrine and their moral beliefs. Applying this framework, however, Shiffrin concludes that US contract law and promissory morality improperly diverge, and she offers the consideration doctrine as a primary example of this divergence. She critiques contract law because promisors have a moral duty to keep their word whether or not there is consideration, yet contract law only regards promises as enforceable when consideration is provided or the promisee has reasonably relied on a promise to his or her detriment.[2]

* I would like to thank Curtis Bridgeman, Richard Bronaugh, Michael Pratt, Song Richardson and Stephen Siegel, as well as participants in a faculty workshop held at DePaul University, for their helpful comments. Any errors are my own.

[1] See S Shiffrin, 'The Divergence of Contract and Promise' (2007) 120 *Harvard Law Review* 708.

[2] *Ibid*, at 709–10.

This article adopts Shiffrin's general accommodationist approach, but challenges her application of that approach to the consideration doctrine.[3] Accommodationists are concerned with the virtuous moral agent's ability to accept the justifications for legal doctrine, given that agent's interest in acting morally. The key legal structure in the contract setting is the private right of action: contract law functions through private rights of action which depend on a promisee's decision to bring suit. This legal structure implicates the morality of the promisee.

Notably, an individual promisee may not be morally justified in coercing the promisor in cases of breach (whether through legal or non-legal means). The virtuous moral agent's understanding of the private right of action is accordingly significant under an accommodationist approach. Moral agents—from the perspective of a promisee—can be expected to see their legal power to initiate litigation as a means to remedy violations of their moral rights, and they may understand the law's justification in these terms. The convergence or divergence of contract law and morality should therefore be assessed with respect to the morality of a promisee's initiation of a private right of action against a promisor.

Given this backdrop, an accommodationist approach requires a determination of when it is morally acceptable for a promisee to seek coercive remedies if a promisor breaks his or her word. This inquiry calls into question Shiffrin's conclusions about the consideration doctrine. In some cases, such as promises for bargained-for consideration, it may be morally acceptable for the promisee to coerce the promisor based on a failure to perform.[4] In other cases, morality will conflict with such coercion. Consideration doctrine is plausibly understood as a legal response to moral concerns about coercive responses to a breach—by refusing to enforce gratuitous promises, courts can avoid enabling promisees to indirectly coerce promisors when morality would conflict with a promisee's efforts to coerce. As a result of differences in the morality of coercive remedies for distinct types of promises, the consideration doctrine may then represent a convergence with promissory morality, and not a troubling divergence. This article suggests that, under reasonable assumptions about promissory morality, the consideration doctrine is actually consistent with accommodationist principles.

[3] As noted, this article focuses on the consideration doctrine. Charles Fried has recently raised doubts about other doctrinal areas in which Shiffrin suggests a divergence exists. See C Fried, 'The Convergence of Contract and Promise' (2007) 120 *Harvard Law Review Forum* 1.

[4] For example, the presence of consideration may justify a claim of ownership of contractual performance by the promisee. See generally A Gold, 'A Property Theory of Contract' (2009) 103 *Northwestern University Law Review* 1 (suggesting a normative connection between the presence of consideration and a promisee's right to insist on contractual performance).

II. THE ACCOMODATIONIST APPROACH TO CONTRACT AND MORALITY

In her recent article, Shiffrin contends that contract law diverges from the morality of promising, and that this divergence is a cause for concern. Before reviewing specific areas of divergence, it is worth considering why such a divergence is relevant. Shiffrin calls for a reassessment of how private law and interpersonal morality are, and should be, related. In her view, a divergence places an unacceptable burden on moral agents who seek both to meet the law's obligations and also to comply with interpersonal morality.

Shiffrin suggests that there are two primary strands of thought on contracts and promises. One approach, which she calls the 'reflective' approach, 'take[s] interpersonal morality as a template for legal rules'.[5] The justifications for a reflective approach vary. Supporters often accept this perspective based on the perceived nature of law, or for reasons grounded in their political philosophy. In either case, this view seeks to move legal content in the direction of morality. An alternative view, the 'separatist' approach, suggests that law and morality should be viewed as independent spheres. Some separatists have 'normative, political reasons' for their opposition to legally forcing people to comply with rules of interpersonal morality.[6] Others simply believe that law, or contract law in particular, has distinct goals from interpersonal morality—for example, the purpose of contract law may be the establishment of an efficient system of exchange.[7]

As Shiffrin indicates, both accounts have advantages to them. However, she concludes that these two approaches do not exhaust the available possibilities. Her insight is to develop another option for assessing the relationship between morality and law, one which borrows insights from reflective and separatist accounts. This approach takes an intermediate, 'accommodationist' position on the proper relationship between morality and law. Shiffrin suggests that 'even if enforcing interpersonal morality is not the proper direct aim of law, the requirements of interpersonal morality may appropriately influence legal content and legal justifications to make adequate room for the development and expression of moral agency'.[8]

Part of the reason for Shiffrin's accommodationist position is her recognition that moral agents believe that 'there are moral duties to obey

[5] Shiffrin, above n 1, at 713.
[6] *Ibid.*
[7] *Ibid.*
[8] *Ibid*, at 715.

the law'.[9] This raises a divergence concern because the law may end up regulating activity which is also subject to distinct moral norms. In light of this possibility, Shiffrin argues that 'legal rules should be sensitive to the demands placed on moral agents so that law-abiding moral agents do not, as a regular matter, face substantial burdens on the development and expression of moral agency'.[10] Contracts, moreover, are a good example of how moral agents may simultaneously participate in legal and moral spheres. Contracts clearly impose significant legal duties, and Shiffrin concludes that promises are embedded in contracts, such that contracts also implicate the moral duties intrinsic to promising.[11]

Based on these premises, Shiffrin derives the following principle:

> [W]hen a legal practice is pervasive and involves simultaneous participation in a moral relationship or practice, the content and normative justification for the legal practice must be acceptable to a reasonable moral agent with a coherent, stable, and unified personality.[12]

In light of this principle, Shiffrin also proposes three criteria against which legal rules should be measured. First, she suggests that legal rules 'should not, as a general matter, be inconsistent with leading a life of at least minimal moral virtue'.[13] Contract law is not typically thought to violate this first criterion. The second criterion is that 'the law and its rationale should be transparent and accessible to the moral agent'.[14] This limitation is more subtle in its operation, and if adopted, its significance for contract law is potentially large. Following this criterion, the justifications proffered for the law's substance should, if accepted, be compatible with the individual's 'developing and maintaining moral virtue'. Moral agents must be able to both understand and accept the law's justification as their own. I will refer to this as the 'transparency' criterion.[15] Lastly, Shiffrin contends that 'the culture and practices facilitated by law should be compatible with a culture that supports morally virtuous character'.[16] This cultural criterion does not require the law to enforce interpersonal morality as such. In fact, legal interference in certain moral areas may be counterproductive. Even so, some legal doctrines, such as doctrines designed to facilitate efficient breaches of contract, might tend to undermine morally virtuous

[9] *Ibid.*
[10] *Ibid.*
[11] *Ibid*, at 721.
[12] *Ibid*, at 717.
[13] *Ibid*, at 718.
[14] *Ibid.*
[15] I follow Michael Pratt's lead in using this terminology. Pratt refers to this as the 'transparency constraint.' See M Pratt, 'Contract: Not Promise' (2008) 35 *Florida State University Law Review* 801, 805.
[16] Shiffrin, above n 1, at 719.

behaviour.[17] To the extent a moral culture is harmed by legal doctrine, this third, cultural criterion comes into play.

The key development in Shiffrin's argument is her suggestion that one need not endorse a reflective approach to legal doctrine in order to be troubled by a divergence between the law and moral principles. One can be deeply sceptical about a legal system that mandates moral behaviour as such, yet still remain concerned with the compatibility between the norms expressed in legal doctrine and the norms commonly accepted by moral agents. Accommodationism allows legal institutions to address this compatibility problem without forcing individuals to comply with any particular comprehensive conception of moral virtue.

I do not intend to challenge Shiffrin's accommodationist perspective in these pages. For the purposes of this article, it will be assumed that her framework is appropriate. Nor will I question her view that promises are embedded in contracts and form their basis.[18] Instead, this article will assess which aspects of promissory morality should be relevant to an accommodationist approach. This analysis, in turn, will cast doubt on Shiffrin's claim that the consideration doctrine diverges from promissory morality.

III. THE PURPORTED DIVERGENCE BETWEEN PROMISSORY MORALITY AND THE CONSIDERATION DOCTRINE

There are several aspects of contract law that Shiffrin offers as examples of a divergence between contract law and morality. These include the consideration doctrine, the use of expectation damages rather than specific performance, mitigation requirements, and the absence of punitive damages.[19] This article leaves most of these potential divergences to one side. Instead, it questions the purported divergence between promissory morality and the consideration doctrine. In the process, this article also suggests that the relationship between promissory morality and contractual *enforcement* is more complex than Shiffrin's argument implies.

As a starting point, it will be helpful to explain the basis of Shiffrin's critique of contract law and the consideration doctrine. She suggests that contract law diverges from the morality of promises because, despite the association of contract obligations with morally binding promises, the

[17] For Shiffrin's argument to this effect, see *ibid*, at 740–43.

[18] *Cf* Pratt, above n 15 (critiquing Shiffrin's comparison of contracts and promises).

[19] Some of these claims are controversial. For further analysis of these other claims, see Fried, above n 3.

resulting legal obligations do not correspond to the moral obligations of promisors.[20] She then uses the consideration doctrine as an example of such a divergence:

> For instance, the moral rules of promise typically require that one keep a unilateral promise, even if nothing is received in exchange. By contrast, contract law only regards as enforceable promises that are exchanged for something or on which the promisee has reasonably relied to her detriment.[21]

Shiffrin's analysis thus focuses on the moral requirement that one keep a promise—that is, it focuses on the moral duty of the promisor to the promisee. From that perspective, the divergence appears to be straightforward, for the promisor has a moral duty to perform regardless of the presence or lack of consideration extending from the promisee. Legal doctrine, in contrast, will often preclude a successful suit if the promise was not made for consideration.[22] From Shiffrin's perspective, promissory morality and contract law differ as to 'whether unilateral promises bind'.[23]

In addition, even if the law does not on its face require citizens to accept immoral premises as justification for legal doctrine, a divergence between contract law and morality could negatively affect the development and maintenance of a moral character. The existing law might encourage individuals 'to associate the conditions of binding agreements with quid pro quo exchange'.[24] Under an accommodationist approach, the existence of legal doctrine that is either premised on ideas which conflict with commonly held views of interpersonal morality, or that is corrosive of moral character, poses a challenge. Shiffrin argues that the consideration doctrine raises these concerns, in light of how moral agents understand contracts.

IV. THE PROBLEM OF PROMISSORY RIGHTS AND REMEDIES

As the above discussion reveals, the content of contract law can be framed to suggest a divergence between law and morality. But it should be noted that Shiffrin's claim is not that we should derive contract law directly from

[20] Shiffrin, above n 1, at 709.
[21] *Ibid*, at 709–10.
[22] Shiffrin also recognises the possibility that distinctively legal grounds may support a divergence between contract and promissory morality. One potential source of distinctively legal grounds, which Shiffrin rejects as a sufficient explanation for the consideration doctrine, would involve the impact of legal enforceability on gift promises. As she notes, there is an argument that gifts do not serve their purpose if they are subject to mandatory enforcement. *Ibid*, at 736–7. For further analysis of this question, see M Eisenberg, 'The World of Contract and the World of Gift' (1997) 85 *California Law Review* 821, 846–52.
[23] Shiffrin, above n 1, at 728.
[24] *Ibid*, at 742.

the morality of promises—she does not endorse a 'reflective' approach. Instead, her thesis builds on 'the law's self-description' that contracts '[rest] upon promises per se'.[25] It calls for a reassessment of contract law based on the extent to which contract law and promissory morality diverge, given the public understanding that contracts are promissory in nature.

Contract theories which purport to explain and justify contract law in terms of promissory morality have been critiqued for failing to distinguish between juridical obligations and merely ethical obligations. For example, consider Charles Fried's theory of contracts. Fried looks to promise-based duties and the value of trust as a basis for contract obligations.[26] This is a reflective approach to the law. However, a reflective approach along these lines also has broad implications. Peter Benson has argued that, in order to explain contract law: 'Fried must hold that a breach of promise, as an abuse of trust, necessarily infringes a right in the promisee that can be coercively enforced.'[27] Fried's theory raises the question whether broken promises necessarily justify coercion.

Accommodationism does not call for contract law to reflect promissory morality in all respects—the role of law under this view is not to enforce promissory morality per se. In addition, accommodationism does not purport to explain how contract law results from that morality. In fact, Shiffrin expressly avoids defining the purposes of contract law.[28] Shiffrin's project is concerned with the proper relationship between contract law and promissory morality. More specifically, she is concerned with the burdens that a virtuous moral agent will face if the justifications for legal doctrine are inconsistent with moral principles. The foundation of this concern is an apparent difference between when promises bind morally and when contracts bind legally, in light of contract law's self-description in promissory terms.

Divergence as applied by Shiffrin has a specific meaning. Shiffrin locates a divergence based on the degree to which the content and structure of contract law 'parallel' interpersonal morality. As she suggests:

> Contract law would run parallel to morality if contract law rendered the same assessments of permissibility and impermissibility as the moral perspective, except that it would replace moral permissibility with legal permissibility and it would use its distinctive tools and techniques to express and reflect those judgments.[29]

[25] *Ibid*, at 722.

[26] See C Fried, *Contract as Promise: A Theory of Contractual Obligation* 17 (Cambridge, Harvard, 1981): 'The moralist of duty thus posits a general obligation to keep promises, of which the obligation of contract will be only a special case—that special case in which certain promises have attained legal as well as moral force. But since a contract is first of all a promise, the contract must be kept because a promise must be kept.'

[27] See P Benson, 'The Idea of a Public Basis of Justification for Contract' (1995) 33 *Osgoode Hall Law Journal* 273, 292.

[28] Shiffrin, above n 1, at 721.

[29] *Ibid*, at 722.

Thus, the absence of a legal duty where a moral duty is present triggers her view that a divergence exists.

The weakness in this theory is in the notion that the structure of contract law should 'parallel' promissory morality. Analogical reasoning is generally subject to the argument that the similarities supporting an analogy may not be the right similarities. Other similarities could always be chosen.[30] Shiffrin's argument falls prey to a comparable difficulty, depending upon which parallels are selected. Parallels between contract law and morality can be found in many ways. Thus, a necessary starting point when finding parallels between contract law and promissory morality is to determine which parallels are the ones which should matter.

Shiffrin's understanding of when contract law parallels the world of promises selects legal and moral permissibility as decisive factors. Yet no guiding principle tells us which features of promissory morality should be the ones which contract law must parallel.[31] If selecting a different parallel between contract law and promissory morality would lead to conflicting results, then the use of parallels will offer limited assistance in analysing contract law. For example, what if the question is, does contract law parallel promissory morality with respect to secondary rights (that is, remedial rights)? A search for parallels here suggests a different outcome from Shiffrin's analysis. In her article, Shiffrin notes that legal remedies and moral remedies may diverge.[32] She appears to view a promisee's moral remedies as a part of promissory morality, and this article will follow the same approach.[33] However, once the promisee's secondary rights are taken into account, it is far from clear that the consideration doctrine improperly diverges from promissory morality.

Although theories of promising vary considerably, it is commonly understood that they involve both rights and duties respecting performance. These primary rights and duties are uncontroversial. The secondary rights which flow from non-performance are much more contentious. A

[30] On the difficulty in determining which similarities should matter in analogical reasoning, see S Brewer, 'Exemplary Reasoning: Semantics, Pragmatics, and the Rational Force of Legal Argument by Analogy' (1996) 109 *Harvard Law Review* 923, 932: noting that 'everything is similar to everything else in an infinite number of ways, and everything is also dissimilar to everything else in an infinite number of ways'.

[31] Although Shiffrin's argument regarding consideration emphasises one parallel in particular, however, Shiffrin does note that various factors would be in parallel if contract law and morality were to 'run parallel'. See Shiffrin, above n 1, at 722 n 24 (suggesting that '[l]egal impermissibility would substitute for moral impermissibility, legal requirement for moral requirement, and so on').

[32] *Ibid*, at 724.

[33] This view of remedies is implicit in Shiffrin's understanding, as she addresses judicial orders of specific performance and monetary damages in terms of a divergence between promise and contract. *Ibid*, at 722–4. If one considers the remedial question distinct from promissory morality, the arguments which follow would remain valid as an accommodationist analysis of moral remedies.

recent article by Margaret Gilbert on promissory obligations provides a good illustration of how these moral secondary rights are often subject to limits.[34] She argues that theories of promising should take into account the view that the promisee possesses a moral right to what was promised.[35] Gilbert does not find, however, that promissory morality permits a promisee to physically force promisors to keep their word. As we will see, this remedial stopping point is significant. Standard intuitions about the morality of coercion can change with context, such as when bargained-for consideration is present. If everyday, gratuitous promises may not be coercively enforced consistent with morality, and if promises with bargained-for consideration may be coercively enforced consistent with morality, then the consideration doctrine offers a parallel between contract law and promissory morality.

A. The Moral Right to Performance of a Promise and its Significance for Moral Remedies

The discussion which follows will use a particular reading of promises to illustrate how a promisee's moral rights can be significant to the accommodationist approach. Gilbert's analysis suggests that the promissory duty to perform need not correspond to enforceable rights held by the promisee. She contends that a promisee has a moral right to performance, yet she questions whether a promisee's morally legitimate responses to breach of an everyday promise include coercive force beyond a rebuke. That said, Gilbert's focus in her discussion of moral rights is on 'everyday' promises. If we turn our attention to other types of promises, coercive remedies become more palatable. Certain promises, such as promises in exchange for consideration, may justify stronger coercive remedies. This part of the article will thus suggest that morally acceptable remedies for promissory breach may fall short of coercion, and in addition, that the acceptability of coercive remedies can vary from one type of promise to another. A comparison of contract law and morality based on the bindingness of contractual promises therefore leaves out a potentially important component—an assessment of the morality of coercive remedies.

The idea of a 'moral right' provides a foundation for the discussion to follow, and it plays a prominent role in Gilbert's analysis of promising. As will become apparent, a 'right' in the contractual setting can have several meanings, covering both moral and legal concepts. The relevant concept in Gilbert's analysis is a moral concept, but in order to understand the idea of

[34] See M Gilbert, 'Scanlon on Promissory Obligation: The Problem of Promisees' Rights' (2004) *Journal of Philosophy* 83.

[35] *Ibid*, at 86.

moral rights, it is helpful to begin with legal understandings. Legal discourse regarding rights covers a variety of related ideas, ranging from claim rights to privileges, powers and immunities, and each are sometimes denoted by the single word 'right'. For clarity, it is helpful to use Wesley Hohfeld's terminology, which differentiates among these legal concepts. Under this rubric, a 'claim right' in the legal setting refers to those rights which correspond to legal duties.[36] If party A has a legal claim right to a particular action by B, then B necessarily has a correlative legal duty to perform this action. This idea of rights in terms of correlative duties is the crucial one for present purposes.

As Herbert Hart has indicated, the correlation between claim rights and duties is more than a feature of law; it has a counterpart in the realm of morality.[37] Drawing on Hart's work, Gilbert recognises the existence of non-legal analogues to legal claim rights.[38] These 'moral rights' involve claims which one party has against another as a moral matter. Corresponding to the right holder's claim, the other party in this moral relationship has a responsibility to provide what is due. If the individual with the duty fails to provide what is due, this constitutes a wrong to the right holder. Promising is an example of a linguistic act which, if successful, ordinarily creates such moral rights. The promisee's moral rights are 'directed'—they correspond to another party's duties.[39] Right holders possess their right against others; the corresponding duty is directed toward the right holder.

Given these premises, there still remains an important point which needs clarification. The presence of a moral right indicates that the possessor of the right should be able to do something about the rights violation. At the most basic level, a right holder should be able to demand conduct consistent with the right. Assuming that promisees have moral rights, one must determine the related question of what, if anything, the promisee can legitimately do to enforce those rights. A common intuition about moral rights holds that where they exist, it is morally legitimate for the right holder to seek to remedy a violation of their requirements.

For Hart, a moral right is linked to the ability to use force. As he suggests:

> The most important common characteristic of this group of moral concepts is that there is no incongruity, but a special congruity in the use of force or the

[36] This distinction is most often associated with the work of Wesley Hohfeld. For Hohfeld's analysis of legal rights, see W Hohfeld, 'Some Fundamental Legal Conceptions as Applied in Judicial Reasoning' (1913) 23 *Yale Law Journal* 16. Gilbert likewise draws on Hohfeld's ideas in this area. See Gilbert, above n 34, at 87.

[37] HLA Hart, 'Are There Any Natural Rights?' *Philosophical Review*, LXIV, 2 (April 1955): 175–91.

[38] See Gilbert, above n 34, at 87 (citing Hart, *ibid*).

[39] *Ibid.*

> threat of force to secure that what is just or fair or someone's right to have done shall in fact be done; for it is in just these circumstances that coercion of another human being is legitimate.[40]

Under Hart's view, when one has a moral right to something, it is legitimate to limit the freedom of another. In some cases, the right holder may even pursue a coercive remedy.[41] Although Hart notes the possibility of legitimate coercion, however, Hart's above description of a moral right does not distinguish between coercion in the form of self-help, and coercion by means of legal mechanisms. The idea of a moral right is potentially consistent with either form of coercion.[42]

Importantly, Gilbert suggests that '[c]ontemplating the exercise of "force" in connection with everyday promises and agreements may seem to go too far'.[43] Although she believes promises create moral rights, her description of these moral rights raises doubts that they are congruent with the use or threat of force beyond a rebuke.[44] Everyday promises as described by Gilbert apparently do not provide the promisee with a full-fledged authority to make others comply by physical force. The promisee has standing to demand performance, or even to rebuke someone who fails in their duties, but more forceful responses are questionable.

Notice, however, that a moral right that supports a rebuke yet falls short of permitting stronger coercive measures is substantially different from a moral right that supports the promisee's use of force to make another perform their promise. Under commonly held intuitions about promising, if someone breaks an ordinary, gratuitous promise, the promisee can publicly demand that he or she perform, and if the promisor fails to do so,

[40] Hart, above n 37, at 178.

[41] Richard Bronaugh has suggested to me that this remedial aspect should be viewed in terms of a promisee's possession of a 'power'. In the legal setting, a good example of a Hohfeldian power would be the promisee's ability to initiate a lawsuit against the promisor. In the non-legal setting, it is an interesting question whether a promisee's response to a breach should be referred to as the exercise of a moral power—say, as to make another liable to moral criticism or release from moral responsibility. For a valuable contribution in this area, see J Raz, 'Voluntary Obligations and Normative Powers' in Supplementary vol XLVI, *Proceedings of the Aristotelian Society* 79, 92–6 (1972) (concluding that the concept of a normative power can be applied outside the law, and developing a theory as to when this occurs).

[42] As will be developed in Part IV.B, both forms of coercion are significant when comparing contract enforcement to promissory morality.

[43] Gilbert, above n 34, at 89.

[44] It should be noted that Gilbert considers a rebuke to fall under the general category of using 'force'. See *ibid*, at 89 ('It is reasonable … to include under the rubric of force the kinds of thing noted at the outset of this essay: such things as informal rebukes and demands.'). I do not share this view. However, whether or not a rebuke should count as 'force', there remains an important distinction between a rebuke and the use of physical force, which Gilbert's discussion highlights. Gilbert's discussion indicates that one may believe that promises create a moral right to performance while doubting that it is moral to remedy a breach of an everyday promise with physical force.

the promisee can rebuke the promisor for dishonesty. It is not usually thought that the promisee may, consistent with morality, personally force the promisor to perform.

This question of whether forcible coercion is morally acceptable is also significant to an accommodationist assessment of contract law. To the extent that contract law is justified in terms of promissory morality, the relevant legal doctrine may be grounded on the moral propriety of coercive remedies as a response to breach, rather than the morality of keeping one's word. In addition, the presence of consideration may be important in determining whether coercive remedies are morally acceptable. Benson's analysis of contract law helpfully suggests this possibility.[45] As Benson indicates, legal doctrines like the consideration doctrine 'suppose a distinction between promises that create correlative rights and duties which are coercible, and promises that may only give rise to an ethical duty of fidelity'.[46] If this distinction exists as a matter of promissory morality, then contract law may parallel what promissory morality requires.

My own writing on contract theory offers support for a moral—and not solely legal—distinction between promises based on bargained-for consideration. I have argued that a contractual promisee can acquire a proprietary interest in a promisor's future actions by meeting the terms of a conditional promise.[47] This view of contracts builds on a Lockean conception of property acquisition. Under this rubric, the doctrine of consideration is a basic feature of an acquisition process—when consideration has been provided, a promisee will understand performance as belonging to him or her. A promisee is able to acquire an enforceable right to contractual performance because the act of providing consideration is analogous to other settings in which an individual's labour justifies enforceable rights of ownership. Providing consideration according to a promisor's terms can thus be a means of obtaining moral rights in performance that are consistent with coercion in cases of breach.

Other theories of promising (and rights acquisition) may also support this conclusion. It is unnecessary for the reader to adopt a particular theory of contracts, or even of promissory morality, to recognise the above distinctions regarding the implications of moral rights. Whatever the source of our moral intuitions about coercive remedies, some promises are reasonably understood to support coercive remedies for breach, and other promises raise serious doubts in this regard.

Scepticism about exercising force—beyond a demand or rebuke—to respond to an ordinary promissory breach is widely shared. This suggests a

[45] See Benson, above n 27. See also P Benson, 'The Unity of Contract Law' in P Benson (ed) *The Theory of Contract Law* (Cambridge, Cambridge University Press, 2001).

[46] Benson, above n 27, at 293.

[47] Gold, above n 4.

range of potentially legitimate responses to a breach of promise. A moral right to performance of a promise could mean the possessor of the right has standing to demand the other party meet the obligation, and to rebuke a failure to perform, or it could mean the possessor of the right has standing to demand performance and also, if performance is not forthcoming, to physically coerce the other party to meet the obligation. These are distinct understandings of moral rights, and the remedies which those moral rights imply.

As shorthand, I will refer to the first example—even if it includes standing to rebuke—as a 'weak' moral right, and the second example as a 'strong' moral right.[48] A weak moral right provides its owner with the standing to demand what is owed, and even to rebuke someone who does not act consistently with their corresponding moral duty. A strong moral right provides its owner with the standing to force compliance with the corresponding moral duty, or to otherwise remedy a failure to perform. Thus, a strong moral right could justify coercion by legal mechanisms, or it could justify coercion by the right holder directly. Weak moral rights do not correspond to remedies that require physical force, while strong moral rights do so.[49]

As will be developed below, this distinction matters when assessing Shiffrin's critique of contract law. If the morality of ordinary promises—promises which lack consideration—does not include a strong moral right of the sort that would support physical coercion as a remedy, then contract law does not diverge from the morality of promising as far as Shiffrin suggests. A judicial unwillingness to legally enforce promises that lack consideration would parallel the moral impermissibility of a promisee using coercion to force performance in the absence of contract law. Furthermore, for contractual promises with bargained-for consideration, the legal power to seek coercive remedies in court may parallel the moral permissibility of using self-help to enforce such promises in the absence of contract law. The selection of parallels between contract law and promissory morality thus has consequences. The remaining question is whether the parallel Shiffrin emphasises—a parallel based on the bindingness of promises—is nevertheless the correct parallel to emphasise given the likely perceptions of virtuous moral agents. The next part of this article addresses that question.

[48] The range of possibilities is actually more elaborate than the two categories described above. A primary moral right could be congruent with a secondary moral right to forcibly coerce performance; a secondary right to forcibly coerce a monetary remedy; a secondary privilege to forcibly coerce performance (or a monetary remedy); a secondary right to rebuke; and other gradations as well.

[49] One might further subdivide moral rights to take into account cases where pursuit of a legal remedy is morally appropriate, yet pursuit of self-help is not. Although that distinction is not made here, it could accurately describe the moral status of contractual promises given the existence of contract law.

B. The Moral Implications of a Private Right of Action

Even if promisees only acquire weak moral rights via ordinary, gratuitous promises, one might still question whether the strength of these moral rights should be meaningful under an accommodationist assessment of contract law. Moral agents (including judges) may not perceive a relationship between the strength of a promisee's moral rights—whether a promisee has standing to coerce when promises are breached—and the law of contracts. There may thus be little reason to focus on parallels between contract law and promissory morality in the remedial context. Would a virtuous moral agent perceive a divergence in terms of bindingness, as Shiffrin does? Alternatively, would a virtuous moral agent perceive a divergence in terms of moral rights? A review of the structure and content of the private law is helpful in analysing these questions.

Shiffrin's divergence claim relies on a comparison of the legal significance of breaching a contract with the moral permissibility of breaching a promise. For Shiffrin, consideration doctrine is problematic because promissory morality requires unilateral promises to be kept, yet contract law 'only regards as enforceable promises that are exchanged for something or on which the promisee has reasonably relied to her detriment'.[50] Implicitly then, Shiffrin is interpreting the non-enforcement of contracts that lack consideration as a legal statement about the moral bindingness of contractual promises. Yet it would oversimplify contract law to say that courts require parties to keep their promises when there is bargained-for consideration, and do not require them to keep their promises in the absence of consideration. What contract law actually does is to provide for the possibility of enforcement contingent upon an individual bringing suit. Courts do not generally enforce legal rights *sua sponte*; they enforce them when a plaintiff initiates litigation. An accommodationist should thus be concerned with the legal and moral justification for private rights of action.

Benjamin Zipursky has recently offered important insights into private rights of action which are relevant here. Zipursky provides an illuminating explanation of how private rights of action, rather than corrective justice or efficiency, are central to an explanation of the private law's function.[51] In the process, his theory also suggests how a claimant's moral right to respond to a breach can be linked to a legal power to bring about enforcement.[52] As will become apparent, this is an area of law where moral

[50] Shiffrin, above n 1, at 710.

[51] See B Zipursky, 'Private Law' in J Coleman and S Shapiro (eds), *The Oxford Handbook of Jurisprudence & Philosophy of Law* (Oxford, Oxford University Press, 2002), 623–55.

[52] Although this right of action is often described as a power, Zipursky has also described it as a Hohfeldian privilege. See B Zipursky, 'Rights, Wrongs, and Recourse in the Law of Torts' (1998) 51 *Vanderbilt Law Review* 1, 81.

agents are confronted with a pervasive legal practice that involves simultaneous participation in a moral relationship.

In Zipursky's view, the private right of action is a key concept in explaining the private law. The use of a private right of action means that the law does not simply enforce corrective justice in cases where one party has wronged another party. Rather than automatically imposing civil liability when such wrongs occur, the private law 'empowers private parties to have other private parties held liable to them, if they choose'.[53] A foundation of liability under this system is the initiation of a lawsuit; the judicial process is optional, and dependent on private choices.

In explaining this structure, Zipursky describes two principles of the private law. The first principle, a right of redress, is drawn from John Locke. Locke famously indicates that there is a natural right to punish wrongdoers. But Locke also indicates that individuals possess a natural right to redress wrongs committed against them. According to Locke, 'he who has received any damages has, besides the right of punishment common to him with other men, a particular right to seek reparation from him that has done it'.[54] This idea of a natural right to redress can readily be found among the norms of the private law. However, Locke also describes the setting up of a judge 'with authority to determine all the controversies and redress the injuries that may happen to any member of the commonwealth'.[55] Under Locke's understanding of the social contract, the state is apparently meant to provide redress for private injuries as a matter of course. As Zipursky suggests, this aspect of Locke's approach does not fit as a description of the private law.[56] The state does not automatically remedy individual injuries due to torts or breaches of contract. This is where William Blackstone's contribution is important.

Blackstone emphasises the private role in determining whether redress will occur. In explaining how the private law works, Blackstone refers to suits in court:

> wherein the act of the parties and the act of law co-operate; the act of the parties being necessary to set the law in motion, and the process of the law being in general the only instrument by which the parties are enabled to procure a certain and adequate redress.[57]

[53] Zipursky, above n 51, at 624.

[54] J Locke, 'The Second Treatise of Government' in D Wootton (ed), *Political Writings of John Locke* (New York, Mentor, 1993) 265–6.

[55] *Ibid*, at 305.

[56] Zipursky, above n 51, at 640 (noting that 'the Lockian picture does not match the law we actually have').

[57] William Blackstone, *Commentaries on the Law of England* (Oxford, Clarendon Press, 1765), bk 3, ch 3, *22.

Locke argues for the right of the individual to redress of injuries, and suggests the state's role in providing such redress. Blackstone notes that the redress provided by the courts is one which is procured by the individual plaintiffs themselves, and consequently the judicial remedy is a redress partly under the individual plaintiff's control.

Zipursky contends that synthesising these two lines of thought suggests 'the view that the power to alter a defendant's legal status through having a judgment entered against him—the private right of action—is something a private party who has been wronged is entitled to from the state'.[58] Moreover, this relationship is not only about an individual's rights, it is also about the state's duties: 'the state—having deprived individuals of other means of self-help—is obligated to empower individuals with an avenue of civil recourse through the courts'.[59] This is a three party normative relationship, one existing between the wronged individual, the state, and the wrongdoer.

From this perspective, private rights of action can be a means of attaining corrective justice, but the normative justification for the state's involvement in providing a remedy is not solely the existence of an unrectified injustice.[60] The state is obligated to provide the right of action in those circumstances where the individual bringing suit would be justly able to seek private redress, but in a civil society lacks that recourse through purely private means. The claimant is still the one initiating the remedial action, but redress occurs through the means of a state-provided institution. In place of non-legal remedies, the claimant is able to exercise a legal power to attain redress through the state.

Importantly, this normative relationship between a plaintiff's legal power to initiate a lawsuit and the state's obligation to provide recourse would be significant to a moral agent such as Shiffrin has in mind. A moral agent will predictably draw connections between the availability of legal recourse and a legal wrong committed by the defendant against the plaintiff. The legal wrong justifies the private right of action. Although Zipursky's explanations for the plaintiff's entitlement to legal recourse are premised on the existence of a legal wrong, however, the right of action also offers redress for a *moral* wrong. Both types of wrong would be relevant to the moral agent. The virtuous moral agent will be interested not only in the legal justification for the existence of a right of action, but also in the

[58] Zipursky, above n 51, at 642.

[59] *Ibid.*

[60] However, it should be noted that a right to redress is not identical to a right to corrective justice. See J Goldberg, 'The Constitutional Status of Tort Law: Due Process and the Right to a Law for the Redress of Wrongs' (2005) 115 *Yale Law Journal* 524, 604 (suggesting that 'redress theory is not tied to a notion of restoration or making whole').

moral question of whether the plaintiff's pursuit of this right of action is a morally acceptable response given the moral wrong at issue.

As Zipursky suggests, 'the explanation of the court's affording … relief [in contract cases] is not simply the defendant's obligation to perform; it is the plaintiff's right to have the defendant perform, in light of the defendant's promise to her'.[61] Zipursky's focus is on wrongs that occur through violation of a plaintiff's legal rights, but his insight is also applicable from a moral perspective.[62] Relief is explicable in light of the plaintiff's moral right to performance based on the defendant's promise to the plaintiff. The option of bringing suit functions as an exception to the general rule that private individuals cannot coerce others.[63] Indirectly, the plaintiff can coerce others through institutional means. A moral agent is likely to see this possibility of legal coercion in terms of redress for a violation of promissory morality—as Shiffrin notes, under US law the contract is represented as an enforceable promise.[64] Thus, in this setting, the moral status of the *promisee* is significant, and not merely the performance obligation of the promisor. A promise which creates a strong moral right legitimises acts of coercion against the promisor (legal or non-legal), while a weak moral right does not.

Admittedly, social contract theories are controversial. What the above analysis does highlight is the likelihood that moral agents will believe that the private right of action is an expression of their moral rights, including the morality of the promisee using coercion against the promisor. Zipursky describes the commonly felt entitlement to a right of redress:

> [I]n light of the fact that we each have instincts to redress wrongs done to us that the state prohibits us from acting upon, and that such a framework of raw liberty to redress wrongs would be of some value to the person who was wronged in terms of self-preservation and self-restoration, the state is obligated to provide someone who has been wronged an avenue of civil recourse, a civil right to redress, through the courts, against the wrongdoer.[65]

The instincts described by Zipursky apply to both moral and legal wrongs. Private rights of action in law often overlap with rights to redress in morality, and for virtuous moral agents, the overlap can be meaningful.

[61] Zipursky, above n 51, at 646.

[62] See, eg, Zipursky, above n 52, at 82 (suggesting tort law can be understood as embodying a principle of civil recourse which 'states that a person ought to be permitted civil recourse against one who has violated her legal rights'). See also Zipursky, above n 51, at 636 ('It is part of the concept of a private right of action as it is deployed in the law that it is always conceived of in relation to a particular act of the defendant who is being sued, where that act is being characterised as to its legal status').

[63] Zipursky, above n 52, at 81; Zipursky, above n 51, at 636.

[64] Shiffrin, above n 1, at 721.

[65] Zipursky, above n 51, at 642.

The idea that certain individuals are, and should be, entitled to pursue private rights of action as a substitute for a moral right of self-help suggests a morality-based structure for the private law. This is especially so in cases involving breaches of contract, where promissory morality is implicated by the wrongs at issue. Private plaintiffs frequently think they should be able to *force* performance of a contractual promise, using courts as a means to accomplish this coercion. This sense of the promisee's entitlement is not just a product of a legal doctrine. If the wrong at issue is seen as a moral wrong, the plaintiff may understand the state-provided right of recourse as a morally legitimate means to redress that wrong, premised on the existence of a strong moral right. Likewise, a virtuous moral agent would also be concerned if the private right of action does not correspond to a strong moral right—in that case a coercive legal remedy would diverge from morality. An accommodationist should therefore be interested in ascertaining whether particular private rights of action converge with commonly accepted moral principles, in terms of the legitimacy of forcible coercion by the claimant. An understanding that the private right of action is a substitute for self-help suggests that the strength of the promisee's moral rights would be a salient issue for a virtuous moral agent, and thus suggests the importance of a parallel between the strength of a promisee's moral rights and the availability of a private right of action.

C. The Significance of the Promisee's Moral Rights for Accommodationism

If a moral agent would believe private rights of action for breach of contract should only be available when they correspond to strong moral rights, this challenges Shiffrin's critique of the consideration doctrine. Shiffrin's claim of divergence would work if gratuitous promises result in strong moral rights, yet it is doubtful that ordinary promises which lack consideration achieve this result—they are often thought to create standing to demand performance or to rebuke a breach, but nothing more. An expansion of contractual liability to cover gratuitous promises on the basis of promissory morality would be likely to create a divergence between the moral agent's understanding of when a promisee may legitimately attempt coercive remedies for breach of promise, and the justification for the state's provision of legal remedies. The state would implicitly be recognising that strong moral rights stem from gratuitous promises when promissory morality would not recognise strong moral rights. That said, the challenge for Shiffrin's critique is potentially more significant. Depending on one's viewpoint, coercion of the promisor in such cases is not only immoral, it is

also an unjust interference with the autonomy of the promisor.[66] Promisors could have a moral right not to be forced to perform a gratuitous promise, even if performance is their ethical duty.

Shiffrin has underscored the overlap between promissory morality and contracts. Contracts, she suggests, involve participating in a legal practice while simultaneously participating in a moral practice.[67] It is for this reason that accommodation of morality becomes important: legally enforceable contracts are more than promises, yet they often implicate the morality of promising. While Shiffrin focuses on the overlap between contractual and promissory obligations, a similar legal and moral overlap exists with respect to remedies. Initiating a private right of action also involves participating in a legal practice while simultaneously participating in a moral practice. The promisee who brings suit to remedy a contract violation is, indirectly, seeking redress for a promissory violation as well.

Once the focus is on moral rights and moral remedies, the ability to use Shiffrin's transparency criterion in this setting is greatly weakened. There is broad agreement that, in general, breaking one's word violates a moral obligation. There is also a widespread belief that promises create standing to demand performance. On the other hand, there is no consensus in support of the view that ordinary promises, lacking bargained-for consideration, create a strong moral right which thereby legitimises the promisee's use of force to make a promisor perform. An absence of consensus on this question means that efforts to comply with the transparency criterion based on the moral views of one subset of the population could easily violate the transparency criterion based on the views of another subset.

The purported divergence found by Shiffrin between contract law and promissory morality with respect to the consideration doctrine depends upon a contested view of promissory rights, and their corresponding remedies. As noted, a leading theory of promising emphasises moral rights to performance yet also questions whether ordinary promises legitimise the use of force (beyond a rebuke) to coerce performance. In addition, contractual consideration can be seen as a way of acquiring strong moral rights, such that forcible coercion will become morally appropriate. In light of the above, consideration doctrine is explicable based on normative distinctions among promises. It precludes legal enforcement when promissory morality would not support coercive measures by the promisee, and it permits legal enforcement when promissory morality arguably does support coercive measures.

[66] It should also be noted that moral concerns in this setting extend beyond questions of moral rights. As Charles Fried notes, there are moral issues that arise in the promissory context that are not part of the morality of promising. See Fried, above n 3, at 8 (discussing an analogy between a promisee and the Good Samaritan).

[67] Shiffrin, above n 1, at 717.

In my view, there are good arguments to support the view that bargained-for consideration supports strong moral rights. It is also commonly believed that ordinary, gratuitous promises do not support strong moral rights. If these positions are correct, the consideration doctrine is unproblematic from the perspective of promissory morality. However, it is not necessary for a particular theory of contracts—or promising—to become the consensus view in order to challenge Shiffrin's argument. A live debate over whether promises for consideration can generate enforceable rights consistent with promissory morality calls into question the use of accommodationism as a means to cast doubt on the consideration doctrine.

Of course, moral claims often raise indeterminacies, and occasional normative disputes do not challenge Shiffrin's theory. But the appeal of her accommodationist approach depends on a commonly accepted set of moral premises, even if it does not require complete agreement on the sources and content of those moral premises. This article suggests the areas of promissory morality at issue for the accommodationist approach are not settled. Significant disagreements as to when a promise creates strong moral rights render it much harder to advocate shifts in legal doctrine based on the burdens the law's justifications will create for virtuous moral agents. Once the focus shifts from a promisor's obligation to perform—a widely accepted duty—to more controversial features of promissory morality, accommodationism no longer provides easy answers.

In response to these challenges, one could assume a particular moral perspective on promising is the correct one. For example, Shiffrin's conclusions could be supported by a view of promising in which the presence of reliance or bargained-for consideration does not affect the morality of coercive remedies for promissory breach. In that case, the import of the accommodationist approach will vary depending on the individual reader's agreement with the chosen conception of promissory morality.[68] Alternatively, an accommodationist could focus only on those moral premises which are widely held. Yet it is questionable whether the assumptions about promising implicit in Shiffrin's article are in fact widely held by persons of virtue, especially with respect to the enforcement of ordinary, gratuitous promises. Without in some way addressing these concerns, accommodationism does not provide an adequate justification for reforming the consideration doctrine.

Moreover, Shiffrin's theory is designed to be consistent with notions that the right should prevail over the good.[69] Her argument is intended to

[68] Taking this path also runs a risk of presupposing a 'particular comprehensive conception of the good or ideal of virtue', a possibility which Shiffrin has attempted to avoid. *Ibid*, at 716.

[69] *Ibid.*

facilitate moral behaviour in those areas where issues of justice underdetermine legal doctrine. The mailbox rule, for example, falls into this category. In contrast to the mailbox rule, the use of coercion to enforce promises does raise justice concerns. If the state is understood as an indirect means for promisees to obtain moral remedies, the existence of the promisee's moral right to pursue those remedies becomes important. Coercion of a promisor to perform or pay damages for breach of contract is plausibly viewed as a violation of the promisor's autonomy, unless there is something more at issue than the mere ethical duty to keep one's word. This concern is legitimate in cases where the promisee only possesses a weak moral right as a result of the promise under dispute.

V. APPLYING THE CULTURAL CRITERION TO THE CONSIDERATION DOCTRINE

Whether or not the transparency criterion justifies eliminating the consideration doctrine, an accommodationist critic of the consideration doctrine might still point to Shiffrin's third criterion—the cultural criterion—for support. This cultural criterion calls into question a legal doctrine if the practical outcome of retaining the doctrine is to negatively affect the maintenance of a moral culture. Here, the focus is on the social impact of a legal doctrine, given that moral agents participate in both legal and moral cultures at the same time.

Indirectly, the consideration doctrine—it is said—can be corrosive of important promissory values. Individuals may come to feel that promises are only really binding in cases where there is a quid pro quo, for example.[70] Moral agents may notice which contracts courts choose to enforce and draw a comparison to those promises which are morally binding. These individuals may then feel that courts do not take promises seriously enough, and this perceived lack of judicial concern could spill over into dealings in non-legal contexts. None of these outcomes can be ruled out as potential social effects of the consideration doctrine.

Assuming that promisors would in fact take their gratuitous promises less seriously as a result of the consideration doctrine, this cultural concern lends force to an accommodationist critique of the consideration doctrine. Granting that this outcome is possible, it is uncertain what overall effects the law of contract has on the morality of promisors and promisees. Does the consideration doctrine make people feel that they don't always have to keep otherwise morally binding promises? This is in part an empirical question. The problem is not solely empirical, however. The question of

[70] *Ibid*, at 742, 752 n 90.

what morality requires must be answered in order to analyse whether a legal doctrine is having a negative impact on moral behaviour.

A complete assessment of the cultural issue again requires analysis of the strength of a promisee's moral rights, and not merely the promisor's primary moral obligations. Implicit in the private right of action—for many, at least—is the idea that the state is acting to permit the use of coercion by legal means which the promisee should otherwise be permitted through acts of self-help. If the doctrine of consideration were removed, the culture of promising could be altered to include the idea that coercion is an intrinsic part of a promissory relationship, available when breaches occur. Promisees could adopt a more truculent attitude towards promisors. This outcome, if it transpired, would not be desirable.

It is ultimately unclear whether more expansive legal understandings of promissory rights would have a broader effect on the general practice of making promises. Educated guesses are possible, but we simply don't know. The risks to our present culture of promising could nevertheless be substantial. As Shiffrin suggests, the practice of promising may be fragile.[71] One legal choice may risk negatively affecting promissory duties, while another legal choice may risk unduly expanding perceptions of a promisee's rights. If Shiffrin is correct, the consideration doctrine risks damage to the public understanding of when promises must be kept, with many individuals feeling less of an obligation to keep their word outside of quid pro quo agreements. On the other hand, the absence of consideration doctrine could increase the readiness of promisees to resort to coercion (legal or otherwise) in order to remedy a breach, in cases where coercion would be immoral.

Shiffrin's cultural criterion offers an important factor to consider in certain contexts. For example, it adds another concern to the debates over the merits of efficient breach theory as a justification for legal doctrine.[72] Notwithstanding this potential relevance to contract law, the effects of the consideration doctrine on a moral culture of promising remain obscure. The empirical question here is not only uncertain, it may actually be unanswerable, even if the attendant moral questions are decisively resolved.[73] Absent stronger reason to suspect the consideration doctrine is damaging to a moral culture of promising, this cultural criterion does not support revisions to the consideration doctrine.

[71] *Ibid*, at 738: expressing concern that 'the culture of trust and promising is fragile in subtle ways that are difficult to track'.

[72] *Ibid*, at 742 (raising the possibility that the law, through its structure or justifications, could encourage individuals to 'engage freely in promissory breach when breach yields only marginal economic gains').

[73] In social science terms, this appears to be a 'trans-scientific' problem. See A Vermeule, *Judging Under Uncertainty* 158 (Cambridge, Harvard University Press, 2006) (describing how a question may be 'trans-scientific' when it is an empirical question, but 'also (in many instances) unresolvable at acceptable cost within any reasonable time frame'.)

VI. CONCLUSION

An accommodationist stance need not require us to jettison the consideration doctrine. An accommodationist stance does suggest the significance of determining when a promisee's moral rights cross the line between a right to demand performance, and a right to force performance or its financial equivalent. Shiffrin's argument works as well as it does because it focuses on an area of promissory morality which is relatively uncontroversial. It is widely accepted that promises create moral duties for promisors to keep their word, and a divergence between law and morality on this issue would naturally be problematic for a virtuous moral agent. Once the focus shifts to the moral remedies available to promisees, however, this apparent consensus proves evanescent.

Accommodationists should not conflate a promisee's right to demand something with a promisee's moral right to coerce the thing demanded. There is a continuum of moral rights at issue, and the effect of a particular fact pattern on a promisee's moral rights should be legally relevant. The consideration doctrine is plausibly related to the morality of a promisee's efforts to obtain coercive remedies under particular factual circumstances. In addition, the legal justifications for a private right of action suggest that a virtuous moral agent would draw a parallel between the availability of legal remedies and the morality of using coercion to remedy a promissory breach. Given these possibilities, the consideration doctrine could readily be consistent with promissory morality.

If a convergence between morality and contract law is sought, these moral questions merit further inquiry. Yet there is no answer under an accommodationist approach as to precisely when promissory morality parallels contract enforcement, unless a particular view of promissory morality is adopted. It is possible to imagine legal doctrines which would be beyond the pale for a virtuous moral agent under almost any conception of the morality of promising. However, such cases are not at stake here. Under reasonable understandings of promissory morality, the consideration doctrine is appropriate. Absent a significant shift in the understandings of virtuous moral agents, the consideration doctrine should thus remain an acceptable feature of contract law, even for an avowed accommodationist.

6

Justifying Damages

CHARLIE WEBB

WHAT CLAIMS SHOULD be available to a claimant who has suffered a breach of contract? The aim of this article is to examine how we may justify particular awards and orders that a court may make in such cases. The focus will be on monetary awards—or damages[1]—though the analysis which follows also bears on other possible orders.

Now, the possible aims of damages awards are at least as broad as the possible aims of contract law generally. In the first instance, our understanding of the incidence and form of damages awards following breaches of contractual obligations will derive from our view of the basis and scope of those obligations (whether or not such awards aim at replicating or enforcing such obligations). Moreover, any argument as to the justifiability of particular damages awards, as with any argument as to the proper content of some part of private law, ultimately rests on a particular conception of the legitimate functions and goals of private law. Though a full account and justification of damages awards must, therefore, depend

[1] The proper scope of the term 'damages' is controversial. On one view, we should reserve this terminology for loss-based monetary awards. The modern tendency, however, is to extend the language of 'damages' to all monetary awards for wrongs, whatever their basis or measure: see, eg, J Edelman, *Gain-Based Damages: Contract, Tort, Equity and Intellectual Property* (Oxford, Hart Publishing, 2002) 5. The issue is particularly confused in relation to contract law, for, while many seem happy to apply the label 'damages' to a variety of awards which serve clearly different purposes and have clearly different bases (eg 'expectation' damages, reliance damages, restitutionary damages), I have not seen anyone refer to claims for payment of an agreed sum (or debt) as claims for damages. This makes sense on the basis that the obligation enforced through such actions is the defendant's primary obligation to pay, an obligation which arises at the point of contract formation rather than later when the contract is breached. As such, these are not monetary awards given *for* breaches of contract. However, contract lawyers do tend to attach the label 'damages' to other awards which similarly seek to effectuate the claimant's primary right to performance: see, eg, the discussion of 'substitutive damages' in the text accompanying nn 58–64. For present purposes, I shall use the term 'damages' to embrace all monetary awards available where a claim is brought following a breach of contract, irrespective of whether the award is, strictly speaking, *for* the wrong of breach.

and be built upon a particular theory of contractual obligation and of private law more generally, I shall not address these questions in any detail. Instead, I shall be focusing on a narrower question: namely, what are the consequences of the recognition of a right to performance when considering how the law should deal with claims brought following breaches of contract? What sort of awards can such a right support?

I should make one point clear at the outset. The approach and arguments developed here are not offered as an account or elucidation of the existing rules on contract damages. Some parts are consistent with the law as it stands, others are not. Rather, the aim of this article is to explore where acceptance of a right to performance takes us when addressing the question of how the law should respond to breaches of contract. In so far as the law deviates from the position taken here, a number of conclusions are possible. One is that the law, though committed to the view that contracts entail rights to performance, also recognises other principles which rightly limit the protection offered to such rights. Another is that, upon closer examination, the law does not, and is right not to, recognise rights to performance as a consequence (or at least as an inevitable consequence) of contract formation. A third possibility is that the law is straightforwardly defective in failing to give adequate recognition to the right to performance when dealing with claims for breach of contract. For now, the key point to note is that the arguments made here are not premised upon the correctness of, and so they do not stand or fall with their ability to fit, the cases.

I. IDENTIFYING THE RIGHT TO PERFORMANCE

That contracts are a source of legal rights and obligations is not in doubt. What is less clear is to what rights and obligations contracts give rise. Those who claim that the law does or should recognise contracting parties as having an interest in, and right to, performance argue that a typical bilateral contract consists of a set of reciprocal undertakings and that a contracting party, upon entering into a contract, acquires a right that her counterpart fulfil or make good her undertakings, while coming under a duty to fulfil her own. So, if I contract to buy your shoes, I acquire a right that you transfer the shoes to me and come under an obligation to pay you the purchase price. You in turn acquire a right that I pay you the agreed price and become subject to an obligation to transfer the shoes to me. This is not uncontroversial. It is not self-evident that entry into a contract should give a claimant a right to the defendant's performance, as opposed to, for instance, a right to the monetary equivalent of performance or a right simply to have compensated losses caused by a failure to perform. Oliver Wendell Holmes famously argued that '[t]he duty to keep a contract

at common law means a prediction that you must pay damages if you do not keep it,—and nothing else'.[2] The existence of a right to performance can and should be questioned.

The scope and content of individual rights and obligations follow from and are determined by the principles that underlie them, and, as such, a full defence of the right to performance would involve an inquiry into why we consider that legal obligations, of any sort, should arise from agreements or mutual undertakings. This would require, at the very least, an article of its own. For present purposes it is enough to point out that a rejection of the right to performance also requires us to reject the notion of breach of contract, at least as conventionally understood.[3] A breach of contract is a failure of the defendant to perform the contract according to its terms. But without a duty to perform, there can be no breach in the defendant's simple failure to adhere to the terms of the contract. A claimant's right to performance is no more than the correlative of the defendant's duty to perform. So, unless we are to say that, whenever we talk of breach of contract, we are really engaged in a sort of fiction or deception, we must accept that there is, in some form, a right to performance.[4] Accordingly, commitment to the notion of breach of contract necessarily involves commitment to the notion of a right to performance.[5]

In any case, I shall proceed on the assumption—shared, I think, by the majority of contract lawyers—that such a right does exist. For those who doubt whether there really is or should be a right to contractual performance, they may understand the argument as taking the form '*if* it is true to say the we have a right to performance, then …'. The validity of the premise must be established, if at all, another time.

[2] OW Holmes, 'The Path of the Law' (1897) 10 *Harvard Law Review* 457, 462. See, to similar effect, OW Holmes, *The Common Law* (Boston, Little, Brown and Co, 1881) 301: 'the only universal consequence of a legally binding promise is, that the law makes the promisor pay damages if the promised event does not come to pass. In every case it leaves him free to break his contract if he chooses.' See too, D Campbell and D Harris, 'In Defence of Breach: A Critique of Restitution and the Performance Interest' (2002) 22 *Legal Studies* 208.

[3] I make a fuller version of this argument in C Webb, 'Performance and Compensation: An Analysis of Contract Damages and Contractual Obligation' (2006) 26 *OJLS* 41, 46–7.

[4] For instance, Holmes' argument requires not only that we deny the existence of any obligation to perform, and correlative right to performance, but also that we reject the view that claims for compensatory damages are founded on the commission of a breach of contract. On this approach, there could be no such thing as a claim *for* breach of contract but only a claim that the defendant fulfil his duty to pay damages in the event of non-performance. See too, Webb, *ibid*, at 45–9.

[5] This leaves unanswered the important question of what performance entails and how we go about determining this: for instance, is the content of the obligation determined by reference to the undertakings the defendant intended to make or to what the claimant (reasonably) understood the defendant to be undertaking?

II. RESPONDING TO BREACHES OF CONTRACT

As we have seen, a breach of contract is committed where a defendant fails to fulfil his contractual undertakings, and hence where the claimant does not receive the performance to which he or she is entitled. A breach of contract may or may not leave a claimant worse off. It may or may not bring a gain to the defendant. A breach of contract which causes the claimant no loss is a breach of contract nonetheless.

So, if in our earlier example you fail to deliver the shoes, you commit a breach of contract. We do not first need to inquire into the value I place on the shoes, and hence on your performance, in order to determine that your failure to deliver them amounts to a breach. So, if I have now decided that I do not like the shoes or have found a better pair elsewhere, while this may well have a bearing on my decision to sue and on what a court might then award me, it does not make your failure to deliver any less a breach. This is just a roundabout way of showing that my right to performance is more than, and so is not reducible to, a right not be left worse off in the event of your failure to perform. I have a right that you perform irrespective of what losses, if any, I may suffer in the event that you fail to do so.[6]

This may be elementary, but it is important not to lose sight of this when we ask how the law should deal with claims brought following breaches of contract. Breaches of contract will often cause the claimant a loss. In such cases, this loss does itself ground a claim and provides a reason for an award.[7] However, as we have seen, the claimant's rights do not stop there. The claimant has a right to performance, and this right is not satisfied simply by ensuring that she is not left worse off as a result of the defendant's failure to perform. This right can be asserted, and should be sufficient to ground a claim, whenever a contract is breached, irrespective of what losses that breach may have caused and independently of any claim that may exist to have those losses made good.

Of course, the most obvious way of giving effect to the claimant's right to performance is by ordering the contract to be specifically performed. This is what we do as a matter of course with obligations to pay money, which are enforced through the action for an agreed sum or debt. However, outside such obligations, the courts order specific performance exceptionally, only where an award of damages is held to be inadequate.

[6] Of course, none of this is to suggest that the law should not *also* recognise the claimant as having a right to have compensated losses caused by the defendant's failure to perform. It should and it does. The point is that the right to performance exists independently of, and cannot be reduced to, any right to avoid or have made good such losses: see further the discussion of Holmes' theory, above n 2 and accompanying text. See too, Webb, above n 3, at 45–9.

[7] See text accompanying nn 42–57.

The conventional view is that damages awards are intended to compensate the claimant for the losses she has suffered as a result of not receiving the performance to which she is entitled.[8] As such, though there remains some uncertainty as to what losses fall to be compensated and how they are to be assessed, the claimant is entitled only to such damages as are necessary to make good any relevant losses. As a corollary of this, where the breach has caused no loss, the claimant can recover no more than nominal damages.[9] However, the rule that specific performance will be available only where such an award is considered inadequate poses a number of problems.[10]

There are undoubtedly good reasons for preferring the simplicity and finality of monetary awards, and so we should be happy to see the courts prioritising damages awards over specific relief where the two provide alternative and otherwise equally effective routes to achieving the same end result. However, if we accept the conventional view of damages awards, specific performance and damages can be seen to serve different purposes. While specific performance is designed to secure for claimants the performance for which they contracted and to which they are entitled, damages seek not to provide claimants with that performance but instead to ensure that they are not left worse off as a result of not receiving that performance. To obtain performance is one thing; to receive a sum of money to make up for the losses caused by not obtaining performance is something different.

[8] See *Robinson v Harman* (1848) 1 Exch 850, 154 ER 363, 365 (Parke B); *Ruxley Electronics and Construction Ltd v Forsyth* [1996] AC 344 (HL) 357 (Lord Jauncey), 360 (Lord Mustill); *Golden Strait Corpn v Nippon Yusen Kubishika Kaisha (The Golden Victory)* [2007] 2 AC 353 (HL) [29] (Lord Scott).

[9] I put to one side claims to gains made by the defendant through breach which may entitle claimants to recover a sum of money in excess of any losses they have suffered.

[10] These problems extend beyond those detailed in the text. For instance, to apply the adequacy of damages test, we must first know what the goal of a damages award is, since only then can we determine whether, on a given set of facts, such an award would do that job adequately. On the orthodox view that damages are compensatory, and given that all losses are necessarily capable of being made good by a monetary payment (see below nn 49–57 and accompanying text), it follows that damages will be inadequate only where the extent of the claimant's losses is uncertain and hence where we do not know what sum to award. However, this does not tally with the cases in which specific performance tends to be ordered; for instance, there is no reason to think that valuing land is any more difficult than valuing other assets, and yet contracts for the sale of land are routinely specifically enforced. See too S Smith, 'Substitutionary Damages' in C Rickett (ed), *Justifying Private Law Remedies* (Oxford, Hart Publishing, 2008) 93, at 105–9. The likely explanation is that courts have been motivated to order specific performance where compensatory damages have been viewed as inadequate *for the purpose of securing the performance to which the claimant is entitled*. However, as argued in the text, securing performance and receiving compensation for the losses caused by non-performance are different things, and, therefore, damages aimed at compensating the claimant for such losses do not seek to secure performance and so will be adequate for this latter purpose only occasionally and never by design. See further, Webb, above n 3, at 51–3.

The adequacy of damages test, therefore, suggests that, when faced with a claim for breach of contract,[11] the default response is not to give claimants the performance they are due but only to make good the losses caused by non-performance. As such, at the end of the day, the claimant is left without that performance but with something else of equal value. This creates an apparent incongruity. At the point of contract formation and when identifying breaches of contract, we regard the claimant as having a right to performance—and the defendant as owing a duty to perform—independently and irrespective of what losses, if any, may result from a failure to perform. However, when it comes to remedying breaches of contract, we appear to disregard the claimant's right to performance and instead are concerned with ensuring only that the claimant is not left worse off as a consequence of not receiving performance.

It looks like something has to give. If the claimant has a right to performance then, where possible, we should see that she obtains that performance. If we consider that the undesirability or impracticability of ordering specific performance is sufficient reason to limit the claimant to compensatory damages, then this calls into question the reality of the claimant's 'right' to performance.[12] Such difficulties are, however, avoided if we can find a way of giving effect to the claimant's right to performance through an award of damages, since we can then have the advantages of a monetary award without denying the claimant the performance to which she is entitled.

III. PERFORMANCE INTEREST DAMAGES

I have argued previously that the right to performance may be capable of being given effect by, and hence may justify, an award of damages assessed by reference to the sum of money it would take for the claimant to purchase from some alternative source the performance the defendant was supposed to provide; in other words, damages assessed on a cost of cure

[11] Or, more precisely, when faced with a breach of a contractual obligation other than to pay a sum of money. As we have seen, where the obligation breached is to pay money, the obligation is routinely specifically enforced through the action for an agreed sum.

[12] The question here is whether we can say that the claimant has a *legal* right to performance, and the defendant a *legal* duty to perform, if the law will in no circumstances require the defendant to perform even when in a position to do so. It is difficult to see why considerations, such as a concern with unduly interfering with the defendant's liberty and the cost and complexity of supervision, which bear on the appropriateness of compelling performance when the courts are faced with a claim for breach of contract, should not equally be factored in when determining whether a right to performance arises in the first place. Either such considerations outweigh the arguments for requiring the defendant to perform or they do not. If they do, we should conclude that the claimant has no right to performance; if they do not, the claimant should have such a right and the courts should give effect to it. See too P Jaffey, *Private Law and Property Claims* (Oxford, Hart Publishing, 2007) 39–45.

basis.[13] I shall only sketch the argument here and address some of the key features of, and potential challenges to, such an award.

As we have noted already, upon entry into a contract the claimant acquires a right to its performance irrespective of the value she places on that performance and the losses she would suffer in the event of non-performance. Traditional compensatory damages awards accordingly fail to give effect to the claimant's right to performance as their purpose is only to ensure that the claimant is not left worse off as a result of breach, rather than to see that the claimant obtains the very performance for which she contracted and which she is due. An award of damages will give effect to the claimant's right to performance only where and to the extent that the award provides the claimant with that performance. However, in so far as such a result can be achieved through an award of damages, we can justify such an award by reference to the claimant's right to performance. The question then is how and in what circumstances a damages award can be said to provide the claimant with performance.

Where the claimant has contracted for some asset or service, it is the defendant's duty to provide the claimant with that asset or service. Where the defendant fails to do this and the court is unwilling to compel the defendant to provide that asset or service herself, it may still be possible for the claimant to purchase an equivalent asset or service, and hence an equivalent performance, from a third party. The claimant's right to performance can, therefore, be given effect, and the defendant's duty to provide that performance enforced, by requiring the defendant to pay the claimant the money needed to obtain an equivalent asset or service from a third party. Albeit by a different route, the claimant ends up with the performance for which she contracted and to which she is entitled.

There are, however, a couple of provisos. Firstly, the claimant's right is to performance and the award of cost of cure damages is premised on the fact that the sum will be used to obtain that performance. Only if the money is so used can it be said that the award gives the claimant the performance for which she contracted. Accordingly, a cost of cure award which is *not* used to secure performance cannot be justified on the basis that it effectuates the claimant's right to performance. In principle, therefore, such awards should be conditional on the claimant actually using that money to obtain the relevant asset or service from another source.[14]

[13] Webb, above n 3. Similar arguments are to be found in B Coote, 'Contract Damages, *Ruxley*, and the Performance Interest' [1997] *CLJ* 537; E McKendrick, 'Breach of Contract and the Meaning of Loss' (1999) 52 *Current Legal Problems* 37; M Eisenberg, 'Actual and Virtual Specific Performance, the Theory of Efficient Breach and the Indifference Principle in Contract Law' (2005) 93 *California Law Review* 975; and, in particular, Smith, above n 10.

[14] Webb, above n 3, at 62–3; Smith, above n 10, at 103. This, of course, contradicts the orthodox principle that what claimants do with any money awarded to them as damages is *res inter alios acta*. This principle makes good sense in relation to compensatory awards,

Secondly, for an award of damages to be justified on the basis that it effectuates the claimant's right to performance, the alternative or substitute 'performance' which the claimant is to purchase from a third party must either exactly replicate the performance to which the claimant is entitled under the contract or differ from that performance only in ways which we can regard as immaterial and so which can be disregarded. So, if I contract with you to do some building work for me and you breach, a cost of cure award can be justified on the basis that it gives me the performance I contracted for only if paying another builder to come in to complete the work can be said to equate (or, at least, be tantamount) to performance of our contract.

Sometimes there will be no difficulty in coming to this conclusion. Wherever a contract allows for the relevant work to be subcontracted out, performance 'through' a third party is clearly possible. However, difficulties arise where the contract provides, and my right is, that *you* do the work. In such cases the performance to which the claimant is entitled comprises not simply an end product but embraces the means by which that end product is provided. How then can we say that the provision of that end product *from an alternative source* equates to performance?

I have argued previously that there are two bases which could enable us to say that, in such cases, a cost of cure award provides the performance the claimant contracted for.[15] The first would be to argue that the contract *does* in fact provide for the defendant paying a third party to provide replacement 'performance' in the event of the defendant herself being unable or unwilling to perform. In most cases this could only be on the basis of the existence of an implied term to that effect, and then the key question is as to the legitimacy of implying such a term. The second would be to say that, though paying a third party to provide the service or asset which the defendant had undertaken to provide herself cannot be said to amount to literal enforcement of the contract (and hence differs from the performance to which the claimant is entitled), it is nonetheless close enough to the performance required under the contract that such an award can be regarded as effectuating the claimant's right to that performance. The issue then is how close is close enough. Certainly some deviation from the exact performance set down in the contract is permitted under the existing rules on the availability of specific relief, where, at the very least, the 'performance' ordered by the court will almost invariably take place at

where the mere receipt of the money makes good the claimant's loss and hence fulfils the purpose of the award (see text accompanying nn 49–57). However, it cannot square with damages awarded on the basis that they effectuate the claimant's right to performance, since here the money does not in itself provide the claimant with performance but only a means of obtaining it. As such, and in contrast to payments made as compensation, we should be concerned with how that money is then used.

[15] Webb, above n 3, at 58–61.

a later date than originally stipulated.[16] Equally clearly, however, there must be a limit to the extent to which we can deviate from or qualify the terms of the contract if we are to be able to justify a damages award on the basis that it provides the claimant with the performance for which she contracted and to which she is entitled.[17]

The argument here marks out a different role for cost of cure awards than the conventional view which sees them as one possible measure of compensatory damages. Indeed, where the claimant has incurred or will incur the expense of remedying the defendant's defective performance or obtaining that performance from a different source, such a sum is potentially recoverable as a loss consequential on the defendant's breach of contract. However, the law places limits on the losses which may be recovered in a compensatory damages claim, and, in particular, it requires that the claimant take reasonable steps to mitigate her losses. This raises problems where the cost of securing performance exceeds the value the claimant puts on that performance.[18] The argument here is that protection

[16] See G Treitel, *Remedies for Breach of Contract* (Oxford, Clarendon Press, 1988) 1.

[17] Smith, above n 10, suggests that the argument for awarding cost of cure damages as a means of securing for claimants the performance to which they are entitled can be extended to claims brought following the commission of torts. Certainly the argument that a damages award can be justified where it gives effect to a right held by the claimant can apply whatever the source of that right and so is capable of extending beyond contractual rights and claims. However, I am doubtful as to how far the primary rights protected through the law of torts can be effectuated in this way. As we have seen, where the performance to which the claimant is entitled is the provision of some service or asset, it will often be possible to acquire an equivalent service or asset, and hence an equivalent performance, via a third party. The primary rights and duties which underlie the law of torts tend to take a different form, however, requiring instead that the defendant refrain from acting in a particular way. Accordingly, though further breaches can be prevented by injunction, in relation to breaches which have already been committed it will usually be too late to secure performance. So, once I have been assaulted or defamed, it is necessarily too late to give effect to my right that you not assault or defame me *on that occasion*. History cannot be rewritten and in no sense can an award of damages, however assessed, be viewed as giving me the performance—that you not assault or defame me—to which I was entitled. All that I can require is that you make good the losses you have caused by the assault or defamation and that you not assault or defame me in the future. Because of this, I suspect that cost of cure awards can be justified on the basis identified in the text only where we can formulate the defendant's primary duty as a duty to secure a certain advantage or end result (even if this is no more than the maintenance of the status quo) for the claimant. This cannot be done in the assault and defamation examples, but it may be possible in relation to, for instance, negligent property damage. Though typically we say that the duty here is to take reasonable care not to damage or destroy another's property, it could easily be reframed as a duty to take reasonable care to ensure that other's property remains undamaged by our actions. This could be broken down into a duty, in the first instance, to take reasonable care to refrain from damaging others' property and, secondly, when it is threatened or damaged by our actions, to take positive steps to (re)secure it (*cf R v Miller* [1983] 2 AC 161 (HL)). So conceived, we can see how it might be possible to say that a cost of cure award enforces that duty and effectuates the claimant's correlative primary right.

[18] See Smith, above n 10, at 110.

of the claimant's performance interest may provide an alternative justification for a cost of cure damages award and which, importantly, is not subject to the same limits as compensatory claims.

The mitigation rule denies the claimant recovery in respect of losses which, though (but for) caused by the defendant, the claimant could reasonably have avoided. In effect, we are saying that it is the claimant rather than the defendant who is primarily responsible for those losses. On this basis, we conclude that they should be borne by the claimant rather than by the defendant.[19] The fact that we have such a rule should not be thought surprising. Compensation claims are necessarily concerned with the question of who should bear certain losses, and it makes good sense for the parties' respective contributions to such losses to be a consideration that the courts take into account when determining by whom those losses should be borne.[20]

Different considerations, however, apply where the claimant is seeking to enforce her right to performance. The claimant is not asking for losses she has suffered to be shifted onto the defendant; the claimant is simply asking for the contract to be performed. If it is objected that a performance interest award enables the claimant to circumvent the mitigation rule and allows her to recover—and requires the defendant to pay—a sum of money in excess of her compensable loss, the answer is that the claimant's rights are not limited to recovery of losses caused by non-performance.[21] The claimant has a right to performance itself, regardless of what value she places on that performance and irrespective of what losses, if any, she may suffer in the event of breach.[22] If it is argued that such an award leaves the claimant with a windfall, then the answer is that this is the very windfall for which the claimant contracted and which the defendant undertook to provide.[23]

[19] Or, at least, we do not have sufficient reason actively to shift the loss onto the defendant, and so allow it to lie where it falls.

[20] Similarly there is good reason for taking into account the claimant's contributory negligence when dealing with compensatory claims (though not in relation to claims which amount to assertions of the claimant's performance interest); *cf Forsikringsaktieselskapet Vesta v Butcher* [1989] AC 852 (HL); *Barclays Bank Plc v Fairclough Building Ltd* [1995] QB 214 (CA).

[21] *Cf Ruxley Electronics and Construction Ltd v Forsyth*, above n 8, at 367, 369 (Lord Lloyd). The same objection can be made against, and similarly fails to defeat, claims for specific relief: see, eg, *White & Carter (Councils) Ltd v McGregor* [1962] AC 413 (HL); *Beswick v Beswick* [1968] AC 58 (HL).

[22] See text accompanying nn 2–6. Nor is the claimant's right to performance conditional on his placing upon it a value which is greater than, or at least proportionate to, the cost to the defendant of providing it.

[23] *Cf Ruxley Electronics and Construction Ltd v Forsyth,* above n 8, at 357–8 (Lord Jauncey). Of course, we are still proceeding on the basis that the claimant is required to use her damages to secure performance from some other source. Without this restriction on the use of the money the windfall objection would often be valid.

In short, if we accept that the claimant has a right to performance, then she should be entitled to a cost of cure award, not because this is the measure of the claimant's compensable loss, but because this gives the claimant the performance to which she is entitled.

IV. THE OUTER LIMITS OF PERFORMANCE

Some obligations once breached cannot then be performed. Goods delivered late cannot then be delivered on time; the confidentiality of information wrongfully disseminated cannot subsequently be restored and preserved. Inevitably the claimant cannot then demand performance of those obligations. At this point it appears that the argument for a damages award which gives effect to the claimant's performance interest runs out, and that, whatever justification there may be for an award of substantial damages, it cannot lie in the protection and effectuation of the claimant's right to performance.[24] In such cases,[25] courts typically award the claimant compensatory damages—that is, damages aiming to make good the loss caused to the claimant by virtue of not receiving the performance to which she is entitled. I shall turn shortly to look at the basis and focus of such awards. Before then, however, I should make clear why we must look beyond the claimant's right to performance if we are to justify such awards.

My argument thus far has been that the claimant's right to performance can support a damages award where and to the extent that this provides the claimant with the performance to which she is entitled. Necessarily, therefore, where performance is no longer possible, there is no scope for awarding damages which seek to secure that performance and which are justified on that basis. Of course, claimants for whom breach has rendered performance unachievable can still complain that their rights have been infringed. Moreover, we can say that any award then made, at least in so far as it seeks to respond to or to correct the consequences of that infringement, is premised upon or grounded in the claimant having a right to performance. However, we need to be able to say why a particular awarded is a justifiable response to that infringement—in other words, why the claimant should be awarded a particular sum rather than another, why *any* sum and not, say, a declaration. In situations where performance

[24] The one qualification is that there may be other obligations under the contract which are still capable of being performed, and so the claimant may still be able to demand performance (whether specifically or through a damages award) of those obligations.

[25] And indeed, as the law stands, in many where performance remains possible.

is impossible, and so where an award does not and cannot provide performance, the claimant's clear right to that performance cannot provide an answer.[26]

Some would argue that this is all a little premature, and that the right to performance can in fact support a damages award even where that performance is now beyond reach. One such argument is founded on the fact that the parties may themselves have provided in the contract for what is to happen in the event of breach. If they have agreed that, upon breach, the defendant is to pay the claimant a particular sum, whether fixed in advance or determined in light of the actual consequences of breach, then an order that the defendant pay that sum to the claimant can be justified on the same basis as the claims we have been examining until now. The award ensures that the claimant receives the sum to which she is, in the circumstances, entitled. Accordingly, though such a claim resembles a conventional award of compensatory damages, it is in fact an example of a performance interest claim, mirroring an action for an agreed sum.

So much is clear. However, on one view contracts *always* make provision for what the parties' respective rights and duties are in the event of breach. If this is true, the availability and scope of *all* claims following a breach of contract can be said to derive from the terms of that contract and hence can be supported by reference to the claimant's entitlement to the performance of that contract. Of course, not all contracts provide expressly for what is to happen should one of the parties breach. Nonetheless, the suggestion is that, even where the express terms are silent on this question, an answer can always be found by identifying a relevant implied term.

What should we make of this? If the claim is that the parties' respective rights and obligations in the event of breach are always to be located in the parties' *agreement*, in the explicit or implicit undertakings they made to one another upon entry into the contract,[27] I find it implausible. Contracting parties will generally, perhaps invariably, be aware that a failure to perform constitutes a legal wrong which can form the basis of a legal claim, and that the court will respond to such a claim by requiring the parties in breach, in some way, to make good their wrong, (usually) through the payment of damages. But I think this still leaves us some way

[26] It should be clear that I am not contending that the question of what rights we have against others should be decoupled from the question of how courts should respond to infringements of these rights. Far from it. Where the factual advantages which the claimant's right seeks to secure can still be achieved, the court's principal response should be to ensure that this happens.

[27] See, eg, A Kramer, 'An Agreement-Centred Approach to Remoteness and Contract Damages' in N Cohen and E McKendrick (eds), *Comparative Remedies for Breach of Contract* (Oxford, Hart Publishing, 2005) 249; *C Czarnikow Ltd v Koufos (The Heron II)* [1966] 2 QB 695 (CA) 730–31 (Diplock LJ); and, most recently, the speeches of Lords Hoffmann and Hope in *Transfield Shipping Inc v Mercator Shipping Inc (The Achilleas)* [2008] UKHL 48.

short of saying that we can always find genuine, though tacit, undertakings covering the parties' responsibilities after breach.[28]

Firstly, awareness that our actions can or will attract legal consequences is not limited to contracting contexts and contractual obligations. I know that punching you on the nose will have various legal repercussions, but we rightly do not view the liability that attaches to me when I do so as deriving from some (implicit) undertaking I have made to that effect. And that is not (just) because of a lack of consideration or intent to create legal relations but, more fundamentally, because *there is no undertaking*. Now, of course, where I walk up and punch you there are no undertakings of any sort, whereas, as between contracting parties, there will be any number of undertakings covering other aspects of their dealings with one another. As such, the search for an undertaking in respect of the consequences of breach makes more sense and is more likely to bear fruit. Nonetheless, I think the point stands: knowing—indeed accepting—certain legal consequences will follow from our actions is not the same as undertaking to do whatever it is the law requires of us.[29]

Secondly, even if we were to say that knowledge and acceptance of contractual rules dealing with the consequences of breach were sufficient

[28] Kramer, *ibid*, at 256–7, appears happy to accept that contracting parties may not intend or undertake to pay damages in the event of breach. Nonetheless, he argues that such awards are grounded in the parties' agreement on the basis that all (primary) obligations are 'oriented to particular consequences' and so the 'risks of such consequences are allocated within the contract'. The implication here is that any given contractual undertaking (and hence any primary contractual obligation) is oriented towards securing for the claimant certain factual advantages, and that these may extend beyond the specific factual advantages which form the express content of that undertaking (since otherwise inquiring into an obligation's 'orientation' would not, as Kramer intends, support liability in respect of consequential losses or indeed justify *any* award where performance of that explicit undertaking is no longer possible). This is no doubt true. Many contracts are made to achieve what may be called ulterior objectives or benefits, that is, benefits beyond, though dependent upon, the simple provision of the promised performance. (For instance, I may want a new crankshaft for my mill so that it can continue to operate and so secure for me the trading profits which would follow from its operation.) The question is where this takes us. Even where the defendant knows what ulterior objectives or benefits the claimant has in mind when contracting, it does not follow that we can say that the defendant *undertook* (or that the claimant can reasonably understand the defendant to have undertaken) to secure such benefits or to compensate the claimant for the losses she may suffer if those benefits are not secured. And, if we cannot identify any such undertaking then we are left looking for some other justification for such an award. Furthermore, often the parties will, reasonably, have different ulterior objectives and benefits in mind. In such cases, even looking at the parties' conduct objectively, there will be no single correct answer to the obligation's orientation. So, I doubt whether an inquiry into an obligation's 'orientation' will often, let alone always, yield a single answer, and, even where we can identify the ulterior benefits the contract is oriented to securing, we need a reason for saying that the defendant should be liable to compensate the claimant for the losses she suffers from being deprived of those benefits.

[29] And similarly, the fact that you know that I am aware of the legal consequences of my actions does not mean that you believe—and does not in itself provide reasonable grounds for belief—that I am undertaking to do what the law requires me to do. *Cf The Heron II*, above n 27, at 730–31 (Diplock LJ).

grounds to infer undertakings to pay damages in the event of breach, I suspect that this would take us only so far. While most contracting parties will know that the law provides for damages to be paid in the event of breach, far fewer are likely to be familiar with the detail of these rules. In such circumstances, the most we can infer is an undertaking to pay such damages as the law requires to be paid. And, if this is true, then looking to the parties' undertakings cannot provide any concrete guidance as to what damages should be awarded, nor can they support any particular measure over others. Once more, we must look beyond the claimant's right to performance to justify the award.

So, inquiring into the parties' agreement or undertakings will sometimes yield an answer to what their respective rights and obligations are in the event of breach, but all too often it will not.[30] Accordingly, the claim that the consequences of breach—and hence the existence and scope of any damages award—can always be attributed to a relevant provision in the contract can only hold true if our inquiry into the terms of the contract does not stop at the parties' express and implied *undertakings*. Of course, we do use the language of implied terms to cover obligations owed by contracting parties which do not derive from their agreement and the undertakings they have given to one another.[31] And, though we typically limit this language to primary terms of the contract, there is no reason why it cannot or should not be extended to obligations imposed on the parties following breach. Indeed, as we have seen, parties are free to provide expressly for their rights and obligations upon breach, and so we could say that, where no such provision has been made, there is necessarily a gap in the contract which can only be filled by the implication of a relevant term.

[30] Accordingly I agree with the basic approach of English courts in cases where the basis or measure of damages is at stake, by which the answer is not thought to be found in examination of the terms of the parties' agreement: see, eg, *Ruxley Electronics and Construction Ltd v Forsyth*, above n 8; *Attorney-General v Blake* [2001] 1 AC 268 (HL); *Alfred McAlpine Construction Ltd v Panatown Ltd* [2001] 1 AC 518 (HL); *cf* the speeches of Lords Hoffmann and Hope in *The Achilleas*, above n 27. See too the statement of Bowen LJ in *Birmingham and District Land Co v London and North Western Railway Co* (1886) 34 Ch D 261 (CA) 274–5: 'a right to damages … is not a right to indemnity as such. It is the converse of such a right. A right to indemnity as such is given by the original bargain between the parties. The right to damages is given in consequence of the breach of the original contract between the parties. It is an incident which the law attaches to the breach of a contract, and is not a provision of the contract itself.' Similarly, I think it is no coincidence that the basic measure of (compensatory) damages in contract is much the same as we find in the law of torts (see, respectively, *Robinson v Harman*, above n 8, at 365 (Parke B) and *Livingstone v Rawyards Coal Co* (1880) 5 App Cas 25 (HL) 39 (Lord Blackburn)). In both cases, the award attempts to put the claimant in as good a position as she would have been in had there been no breach, but, typically, in the latter case, there will be no agreement or undertaking between the parties from which such an award might be said to derive.

[31] *Luxor (Eastbourne) Ltd v Cooper* [1941] AC 108 (HL) 137 (Lord Wright); *Shell UK Ltd v Lostock Garage Ltd* [1976] 1 WLR 1187 (CA) 1196 (Lord Denning MR); E Peel, *Treitel: The Law of Contract*, 12th edn (London, Sweet & Maxwell, 2007) 223, 229–36.

If this is true then it seems that *any* damages award can be regarded as provided for by, and so as giving effect to, an implied term of the contract. This in turn appears to open up the possibility of supporting *all* such awards by reference to the claimant's right that the contract be performed to its terms. Now, this move is in itself unobjectionable, but its effect is to bring within the single notion of a right to performance a collection of rights which have diverse bases and justifications. Some terms of the contract, and the rights and duties to which they give rise, derive from the undertakings the parties have themselves given. Others terms, and so other rights and duties, arise, despite the fact that the parties have not given any such undertaking, because some other principle of justice or efficiency or the like is thought to justify it. I have been using the language of the right to performance to cover only rights falling within the former class. No doubt we can extend the terminology of performance so that the latter set of terms, and correlative rights and duties, is also embraced.[32] But the difference between them remains. We just have to find different language to express it.

My claim is that, whatever view we take of the basis of contractual obligation, if we recognise that contracts arise out of undertakings and lead to obligations to fulfil or make good those undertakings, the principle which supports such obligations (and their correlative rights) cannot account for whatever additional obligations the law recognises as owed by and to contracting parties that cannot be said to embody or to derive from any such undertaking. Where the parties have made no undertakings as to their rights and obligations upon breach, and performance of the undertakings they did make is no longer possible, any damages award cannot be justified by reference to the undertakings the defendant did make and the principle which justifies the obligation to make good those undertakings.

This brings us to a second challenge to the position taken here, that the right to performance can support a damages award only where and to the extent that it provides the claimant with that performance, and hence that we must look elsewhere to justify damages awards where performance is no longer possible. The objection here is one recently expressed by Ernest Weinrib: that although the specific action required of the defendant may change after breach—for example, to pay compensation rather than to deliver the shoes she undertook to provide—the duty, and its correlative right, remain the same. As such, we should say that, although the

[32] Though, so understood, the claimant's right to performance becomes the right that the defendant fulfil whatever obligations the defendant owes to her, or, pleonastically, the right to have one's rights effectuated. As such, the notion of a right to performance becomes essentially meaningless.

'performance' to which the claimant is entitled may change, it is the claimant's right to performance which is effectuated through, and so which supports, the damages award.[33]

Why should it be thought better to see these as two manifestations of the same right, rather than simply different rights? For Weinrib this follows from a proper understanding of the nature of private law claims and, in particular, of the relationship between rights and remedies. Weinrib argues that a claimant's rights, and their infringement by the defendant, are not simply conditions of the claimant's claim and any remedy she is given, but are the *reason* for granting a particular remedy.[34] The injustice of which the claimant complains when she brings a claim is that some right of hers has been infringed, and the point of the remedy—the order or award made by the court—is to correct that injustice by realising or effectuating the claimant's right. On any other view, we must say that the injustice suffered by the claimant, while a condition for granting some remedy, does not dictate the content of that remedy. Yet, if this is so, we have no reason for making the availability of any given remedy dependent on the identification of a particular injustice.[35]

As I hope is clear already, I agree with much of this. I certainly believe that when faced with private law claims courts should be seeking to identify the rights of the claimant and, where possible, to give effect to those rights. Indeed, I would say that where there is no right of the claimant to which the court can give effect, there is no basis for any award or order in favour of the claimant beyond a declaration that her rights have been infringed. Nonetheless, I see no reason to accept Weinrib's argument that, at least where 'exact' performance is no longer possible, breach changes the content of the right to performance to a right to compensatory damages, such that the right to performance lives on in and is effectuated by such an award.

Firstly, Weinrib appears to assume that to reject his understanding of such awards involves a disconnection of right and remedy and so leads to the sort of incoherence described above. This would indeed be true if the award could not be said to realise some right of the claimant. But this is not what I am suggesting. My claim is simply that, where performance is no longer possible, a damages award cannot give effect to, and so be supported by, the claimant's right *to performance*. A damages award may still be justified if we can identify some other right of the claimant's, one which is still capable of being effectuated by such an award. This is, I suggest, the case with claims for compensatory damages. As such, there is

[33] E Weinrib, 'Two Conceptions of Remedies' in C Rickett (ed), *Justifying Private Law Remedies* (Oxford, Hart Publishing, 2008) 3.

[34] *Ibid*, at 15, 26.

[35] *Ibid*, at 7–8.

no reason for thinking that the logic or structure of private law demands that we view the right effectuated through compensatory awards as simply another manifestation of a claimant's right to performance.

The question then is whether Weinrib's is a better or more compelling account of a claimant's rights than the one I am offering here. I have argued that a given right can be said to be effectuated by, and so can itself support, a particular award or order only where that order or award provides the claimant with the very factual advantages which that right seeks to secure for her. Only then does the award give the claimant the very thing or performance to which he or she is entitled. Weinrib, however, suggests that there are in fact two ways in which (primary) rights, such as the right to performance, can be effectuated or 'restored'.[36] Firstly, and more straightforwardly, there is what Weinrib terms 'qualitative restoration'. These are the sorts of awards I have been describing, whereby the claimant is provided with the exact thing or performance she is entitled to, as exemplified by orders of specific relief. The performance interest damages awards I have argued for above would also fit in here. Secondly, Weinrib argues that a right may be effectuated in its 'quantitative form'. Here, the claimant obtains not performance itself but the monetary value of that performance. This is exemplified by standard awards of compensatory damages.

For Weinrib then, 'quantitative restoration', though giving the claimant something different from her initial right (or, at least, from her right as initially formulated), is nonetheless to be regarded as a realisation of that right. Why? It is of course true to say that we can put a value on rights and the factual advantages they aim to secure for us. Accordingly, where a right has been infringed and is not or cannot be specifically effectuated, we can still ensure that the claimant receives something else of equal value to the performance of which she has been deprived. Such a claimant will then not be left worse off because of the breach. But this is not performance. As I argued earlier,[37] a claimant, upon entry into the contract, acquires a right to performance irrespective of the value she places on that performance. The defendant's duty is not to provide the claimant with something of equal value to the performance for which the claimant contracted, nor is it simply to ensure that the claimant is not left worse off by virtue of not receiving that performance. It is a duty to perform, to provide the performance the defendant undertook to provide. So, in so far as Weinrib appears to be suggesting that there are two aspects or forms of the right—the qualitative aspect or form, focusing on the specific factual advantages which the claimant can demand the defendant provide, and the

[36] *Ibid*, at 13.
[37] See text accompanying nn 2–6.

quantitative, which looks to the value to the claimant of those advantages—I think he is wrong.[38] If a defendant breaches a contract or commits a tort which happens not to leave the claimant worse off, we do not view this as only a partial breach, as a breach only of the 'qualitative' aspect of the claimant's right but not the 'quantitative'. There are not two strands to the claimant's right. There is one indivisible performance to which the claimant is entitled and which the defendant is under a duty to provide.[39]

Weinrib's answer would appear to be that, though this is all true when the right first arises, breach changes things.[40] Breach changes, or at least

[38] I also think that Weinrib's account has a difficulty in accommodating damages for consequential losses, that is, losses consequent upon and so additional to the loss inherent in the defendant's defective performance. Even if we do say that a claimant's right to performance has a quantitative form or can be viewed in quantitative terms, and hence that a damages award can be said to realise the claimant's right to performance so understood, this would seem only to explain recovery of the value of that performance (or the difference between the value of that performance and the defective 'performance' in fact provided by the defendant). The claimant's right to performance is a right only to the specific factual advantages which the defendant undertook to provide. Any further (consequential) advantages which the claimant may then derive from performance are not part of that performance and hence are not part of, or protected by, the claimant's right to performance. For example, say, I want a new crankshaft for my mill, and agree to buy one from you. Even though that crankshaft will, or may, enable me to make certain consequential gains, my right to performance is simply a right that you provide me with that crankshaft. Accordingly, if my right to performance is to be assessed or understood quantitatively, it can cover only the value of that crankshaft. If I am also to be able to recover the additional, consequential losses I suffer as a result of not receiving the crankshaft, it cannot be on the basis that these are part of the performance to which I am entitled.

[39] To similar effect, Arthur Ripstein (A Ripstein, 'As if it Had Never Happened' (2007) 48 *William and Mary Law Review* 1957) has claimed that compensatory damage awards effectuate a claimant's primary right on the basis that such awards ensure that the claimant retains the 'means' (meaning the things or attributes—such as property and physical and mental powers—which are yours to employ in the pursuit of your 'ends') which the primary right sought to secure for her. Awarding a sum of money to a claimant amounts to a provision of such 'means'. An appropriate sum will ensure that despite, for instance, being deprived of the shoes she was promised or despite the physical injuries the defendant caused her, the claimant is left with no 'fewer' means—ie, that she is in as good a position to pursue her ends (ambitions, goals etc). But, as with Weinrib's notion of quantitative restoration, this can be regarded as effectuating the claimant's primary right only by stripping it of its specific content. So, rather than being understood as a right to a pair of shoes or to physical integrity etc, every primary right is treated as a right to the shoes/physical integrity *or* equivalent means. On this view, 'means' are essentially fungible. But when looking at contractual obligations, this does not reflect the content of the undertakings which we regard as generating those duties. Typically, my undertaking is to provide you with the shoes, not to provide you with the shoes *or* any equivalent sum of money. Indeed, I think the same is true of primary obligations throughout private law. They are not obligations to secure some abstract measure of means, but to provide us with certain specific factual advantages or to avoid certain specific factual harms. This also shows that Weinrib and Ripstein's views are not far removed from (and are open to many of the same criticisms as) the position taken by Holmes (see discussion above, nn 2–4).

[40] Weinrib, above n 33, at 26.

can change, the content of the claimant's right, and hence the 'performance' to which she is entitled. But this appears impossible to square with any view of contract which sees the parties' obligations as deriving from and defined by reference to the undertakings they have made to one another. If (or to the extent that) we believe that contracts arise from agreements or undertakings and involve obligations to fulfil those undertakings, the scope of those obligations will be determined, at least in the first instance, by the scope of those undertakings. So, where we contract for you to sell me your shoes, my obligation to pay you derives from that agreement, from my undertaking. The same goes for your obligation to deliver the shoes to me. But, unless you have explicitly or implicitly undertaken this, the same cannot be said of any obligation you come under to pay me compensatory damages should you fail to give me the shoes. To account for this obligation, and its correlative right, we must look beyond whatever principle it is that supports holding us to the undertakings we have made. So, though Weinrib argues that compensatory awards realise a claimant's right to performance, despite their different content, because they have a common 'normative ground',[41] in fact it is by looking to the respective justificatory bases that we can best see why compensatory damages awards *should not* be seen simply as manifestations of a claimant's the right to performance.

Of course, if this is correct, it means that where performance is no longer possible the claimant's right to performance goes unrealised. One might contend that Weinrib's argument, therefore, has the advantage of not forcing us to concede that, sometimes, rights will go unprotected and be left uneffectuated. But, on this view, the right to performance is given effect only by stripping it not only of its content but also what may be considered its defining feature—its basis in the undertakings the defendant has made to the claimant.

V. COMPENSATORY DAMAGES

This brings us at long last to awards of compensatory damages. As I have argued over the preceding pages, these awards are best regarded as manifesting and effectuating a right of the claimant's which is distinct from her right to performance. This right is commonly described as a 'substitute' for the claimant's right to performance,[42] but I do not think this quite captures the nature of the right or its relationship to the right to

[41] *Ibid*.

[42] See, eg, *Photo Production Ltd v Securicor Transport Ltd* [1980] AC 827 (HL) 848–9 (Lord Diplock); D Friedmann, 'The Performance Interest in Contract Damages' (1995) 111 *LQR* 628, 629–31. For other uses of the language of substitute or substitutionary rights: see below nn 58–64 and accompanying text.

performance. The language of substitution suggests that, upon breach, the right to compensation for losses caused by the defendant's failure to perform arises in place of the claimant's right to performance. However, even where performance has been put beyond reach, I see no reason for concluding that the claimant's *right* to that performance has been lost.[43] Certainly where performance remains possible we should continue to regard the claimant as entitled to that performance and our principal response to claims for breach of contract should be to ensure that this performance is provided. Moreover, claimants who have suffered a breach of contract (and courts faced with the task of responding to such claims) need not choose between demanding that the contract be performed and calling on the defendant to make good the losses caused by non-performance. Hence, a claimant should be able to demand both that the contract be performed, *and* that any losses which performance would not correct, for instance those caused by delay, be compensated.[44] The only limit should be the usual rule that we do not allow double or inconsistent recovery. As such, I think it is preferable to view the right to compensation, not as a substitute or replacement for the right to performance, but simply as a distinct right with a distinct objective, and which exists and may be asserted alongside the right to performance.

What then is the content and basis of this right? Here I plan to say very little on why contracting parties should have, in addition to a right to performance, a right to have losses caused by non-performance made good. As I have argued already,[45] not only is such a right distinct from the right to performance, but I think it must also be seen as resting on a distinct principle. My inclination is that, whatever this is, it is the same principle which supports compensatory awards for torts and other wrongs. For present purposes, however, I shall leave this question to one side and, instead, take a closer look at the loss which compensatory awards seek to make good.

The language of 'loss' tends to be used in two different, though related, senses. Firstly, we sometimes use the word 'loss' to describe certain specific physical harms or unwanted occurrences. In this way, a broken leg, smashed window or an undelivered pair of shoes can all be described as losses. Understood in this way, loss is inherent to breaches of contract, indeed to all breaches of duty. Every time there is a breach of contract, the claimant ends up being deprived, if only temporarily, of some part of the

[43] See text accompanying nn 60–61.

[44] This indeed is the position the law currently takes whereby a claim for specific performance or an action for an agreed sum can be combined with a claim for compensatory damages: see, eg, s 49 of the Supreme Court Act 1981; A Burrows, *Remedies for Torts and Breach of Contract*, 3rd edn (Oxford, Oxford University Press, 2004) 506.

[45] See text accompanying n 32.

performance, of something—an asset, service and so on—to which she is entitled. In simple terms, breach means that, as a matter of factual reality, the claimant does not get something she should have got.[46]

Now, if were are to understand 'loss' as specific physical harms or occurrences, losses can be made good, if at all, only by ensuring that the relevant harmful or unwanted state of affairs no longer obtains. So if the loss of which the claimant is complaining *is* a broken window or not receiving a pair of shoes, then that loss can be corrected only by seeing that the window is repaired and so left 'unbroken' or by ensuring that the shoes are delivered. When loss is understood in this way, a claim to have that loss made good is simply another way of describing a claim that the contract should be performed and hence that the claimant should be provided with the very factual advantages which the defendant undertook to provide and to which she is entitled. As such, these 'losses' are reversed by orders of specific relief and, as suggested earlier, awards of damages which enable the claimant to obtain performance from some third party. It also follows that where performance is no longer possible, and hence where the specific factual advantages which performance would provide can no longer be obtained, then the loss of those advantages can no longer be corrected and hence cannot form the basis of a claim.

There is, however, a second use of the word 'loss'. This describes not the factual, physical harms or adverse occurrences which we may suffer but, broadly, the impact they have on us, on our happiness or quality of life. In this sense we speak of the broken leg, smashed window or undelivered shoes as a *source of loss* or of the loss *flowing from* such occurrences. Since loss here is not identified as specific physical harms we may suffer but is instead something consequent upon them, its reversal is not dependent upon and so need not involve the reversal of those specific physical harms, and indeed such a reversal is possible even where those harms can no longer be made good. Rather, if loss is conceived as a worsening of the claimant's position, a reduction in her quality of life, then making good that loss requires that we do something for the claimant which improves her quality of life sufficiently to offset the reduction brought about by the defendant's breach. It is loss in this sense which I believe is the proper focus of compensatory awards and which provides the basis of an award of damages even where performance of the defendant's primary obligations is no longer possible.

Though losses will always result from real world occurrences, the loss itself is distinct from those events. We suffer a loss every time something

[46] This is not quite the same as saying that the breach of contract is itself the loss, since, on this understanding, loss describes the factual impact on the claimant of the defendant's conduct (eg, not receiving the promised pair of shoes, an insufficiently deep swimming pool), whereas breach describes the legal 'quality' of that conduct.

happens—we experience or receive something—which we value less than the state of affairs which existed before or which would otherwise have come about (depending on our point of comparison). So, when a claimant who has suffered a breach of contract complains of having suffered a loss in this sense, her complaint is not (simply) that she has not received the factual advantages which performance would have brought but that, because she has been deprived of those advantages, she is now in a worse position—one she values less—than that which she would have occupied had performance been forthcoming. Here then loss is not an inevitable feature or consequence of a breach of contract. A claimant will suffer a loss only where and to the extent that she values performance and the factual advantages it would bring her more than the defective 'performance' which the defendant in fact provided.

This has a number of important consequences. Firstly, the measure, and indeed the very existence, of a loss follow from and will be determined by our preferences and objectives. Because these vary from person to person, loss is necessarily personal or subjective. What I regard as a loss is likely to be different from what you consider to be a loss since we have different tastes and goals and hence value things differently. For this reason the courts are correct to draw attention to the 'consumer surplus'; the possibility that, and the extent to which, a claimant places a value on performance in excess of its market value.[47]

Secondly, loss is not co-extensive with, and so does not depend on, a reduction in the funds or assets at our disposal. We clearly can and do value assets, and a loss of assets is one possible source of loss. However, it is clear that there are other 'things' which can also make a contribution to our ability to satisfy our wants and fulfil our objectives, and hence to which we also attach value, such as experiences, knowledge and relationships. It is because of this that we do not find it counter-intuitive that people choose to pay money for holidays or to see a film or even make donations to charity. By making such choices, we are saying that we value the experience of the holiday or film, or the welfare of the beneficiaries of the charity, more highly than the money itself (and the other 'things' which that money could provide for us). By contrast, if we were to consider loss

[47] See, eg, the speech of Lord Mustill in *Ruxley Electronics and Construction Ltd v Forsyth*, above n 8 and the speeches of Lords Scott and Hutton in *Farley v Skinner* [2001] 3 WLR 899 (HL); see, too, D Harris, A Ogus and J Phillips, 'Contract Remedies and the Consumer Surplus' (1979) 95 *LQR* 581. One might even suggest that the language of the consumer surplus is misleading both in that it suggests that, when measuring loss, our starting point should be market value (with the consumer surplus being a deviation from and exception to this measure), and that the relevance of subjective value is limited to consumer contracts. As a matter of principle, loss is always subjective and market value is only ever an indication of what the claimant's (subjective) loss may be. In practice, however, given the difficulty of establishing the value the claimant places on performance, it is understandable that courts should use market value as their benchmark.

as dependent upon and co-extensive with a reduction in our assets or wealth, we would be forced to conclude that all contracts for the provision of a service, or any other contract which does not leave the claimant with some 'marketable residuum',[48] necessarily leave the claimant worse off. A breach of such a contract could cause a loss to the claimant only if (and then only to the extent that) she paid in advance or parted with other assets in reliance on the defendant's performance. Similarly, such a restrictive understanding of 'loss' leaves no room for recovery in respect of personal injuries or illness, save in so far as these lead us to incur out of pocket expenses.

More fundamentally, we can challenge the preconception that losses stemming from a reduction in one's assets are of a different order or kind from those flowing from, for example, a lousy holiday, disrupted night's sleep or twisted ankle.[49] If we ask why it is that we attach value to assets, the answer is that, through their use, we are (or may be) better able to pursue our goals and satisfy our wants. This may seem to contradict the view that an asset's value is determined by what others are willing to pay for it. However, not only is sale among the 'uses' to which one can put one's assets, but, more importantly, the reason we are all willing to pay for assets is precisely because we value the uses to which those assets can then be put. As such, market value follows from and is a reflection of the value people generally attach to use and enjoyment of the asset. An asset which is no use to anyone will have no market value.

Accordingly, we value assets in the same way and for the same reason that we value experiences, knowledge, relationships and the like; because of the contributions they make to our quality of life, to our ability to achieve our goals and satisfy our wants. Consequently a reduction of our assets is a loss precisely because it inhibits the pursuit of those goals and so reduces our quality of life. So, while we can certainly distinguish losses in the sense of specific factual harms—property damages, physical harm, psychological injury and so on—once we understand loss as something consequent upon, and hence distinct from, those harms, we must understand such losses as necessarily fungible.

[48] This expression is used in unjust enrichment literature to distinguish services which leave the recipient with some increase in the assets or wealth at her disposal, and 'pure' services which do not: see J Beatson, 'Benefit, Reliance, and the Structure of Unjust Enrichment' in *The Use and Abuse of Unjust Enrichment* (Oxford, Clarendon Press, 1991) 29–32. The understanding of 'loss' put forward here is in substance the mirror image of the notion of 'benefit' or 'enrichment' employed by those who argue that pure services can be enriching.

[49] As manifested in the distinction commonly drawn between 'financial,' 'economic' or 'pecuniary' losses and 'non-pecuniary' losses: see, eg, Burrows, above n 44, at 29; P Cane, *The Anatomy of Tort Law* (Oxford, Hart Publishing, 1997) 98–9; *Wright v British Railways Board* [1983] 2 AC 773 (HL) 777 (Lord Diplock); *Heil v Rankin* [2001] QB 272 (CA) 293 (Lord Woolf MR).

This may all sound very different from the way the notion of 'loss' is typically presented in the cases and texts. Nonetheless, I do not think there exist any great differences of substance between those accounts and my own. In particular, the fact that we award compensatory damages in cases where performance is no longer possible, and hence where the relevant factual physical harm suffered by the claimant cannot be put right, shows that we do not equate loss with such factual harms. It is clear, however, that the courts have, on occasions, lost sight of the fact that losses can occur other than through a diminution of or damage to the assets at the claimant's disposal.

Take, for instance, situations in which A contracts with B for some asset or service to be provided to C. Here, the problem has been seen to be that, if B breaches, it is C and not A who suffers the loss, since it was C and not A to whom the asset or service was to be supplied. This then appears to lead to the result that A is unable to recover anything more than nominal damages, which, when combined with the common law rule that C, who *had* suffered a loss, had no claim, meant that the law let breaches of such contracts go largely unremedied.

In fact, we have two clear bases for awarding A substantial damages in such cases.[50] Firstly, A has a right to performance of that contract, and so, as I argued earlier, we can justify awarding her damages assessed on a cost of cure basis so as to enable her to purchase an equivalent 'performance' from another source. Secondly, and alternatively, A can say that *she* has suffered a loss as a result of B's breach of contract. The fact that A was prepared to pay B to provide the relevant service or asset to C, shows that A placed a value on that performance. Where that performance is not forthcoming, A is accordingly deprived of something she values. The defendant's breach has led to a state of affairs—C not receiving the relevant asset or service—which A values less than the state of affairs which should have eventuated—proper performance and C's receipt of the intended asset or service. This is a loss. We are led to the conclusion that there is no loss here only by adopting a conception of loss which looks only to the quantum and value of the assets at the claimant's disposal. As we have seen, such a restricted notion of loss is insupportable.

Another example of the courts making heavy weather of 'loss' is provided by the line of cases exemplified by *Wrotham Park v Parkside*

[50] A failure to distinguish between these two alternative bases for an award of damages is evident in the various judicial attempts to devise a principled solution to the problem posed by these cases. See, eg, the dicta of Lord Scarman in *Woodar Investment Development Ltd v Wimpey Construction UK Ltd* [1980] 1 WLR 277 (HL) 300–1 and Lord Griffiths in *Linden Gardens Trust Ltd v Lenesta Sludge Disposals Ltd* [1994] 1 AC 85 (HL) 97, and the approaches favoured by Lords Goff, Browne-Wilkinson and Millett in *McAlpine v Panatown*, above n 30. See generally Webb, above n 3, at 53–6.

Homes.[51] Here the defendant built on its land in breach of a restrictive covenant. The claimant, to whom the obligation was owed, sought an injunction to restrain further building and for the demolition of the buildings already erected. Brightman J, however, refused to order their demolition on the basis that this would be a waste of much needed housing. The claimant was, however, entitled to damages and, though the defendant's breach had not affected the value of the claimant's land, Brightman J awarded the claimant damages of £2,500, on the basis that this was the sum of money which 'might reasonably have been demanded by the [claimant] from [the defendant] as a quid pro quo for releasing the covenant'.[52]

It seems tolerably clear that Brightman J considered this sum compensation for a loss caused by the defendant's breach of duty. However, this understanding of the award has come in for considerable criticism over recent years. The supposed flaw in the compensatory analysis of cases such as *Wrotham Park* is that the claimant in fact suffers no loss, and, as such, there is nothing to compensate. Focusing on Brightman J's reference to the fee that the claimant could have demanded to release the defendant from its obligation,[53] the standard objection is that this can only be viewed as a genuine loss where the claimant would in fact have been willing and able to obtain such a sum in return for granting a release to the defendant. Where, as in *Wrotham Park*,[54] no such release would have been granted, any such 'loss' is routinely dismissed as a fiction.[55]

No doubt it is a fiction to say that the claimant has lost a sum of money which it would not in any case have obtained. Nonetheless, we should have no difficulty identifying a loss on such facts. The claimant did not want the defendant to build on its land. When the defendant, in breach of the obligation it owed to the claimant, did build, it brought about a state of affairs which the claimant wished to avoid, a state of affairs which it considered worse than the state of affairs which obtained previously and would have persisted had the defendant abided by the restrictive covenant. This is a plain loss. The fact that the claimant would not have released the defendant from its obligation, even in return for a payment of thousands of pounds, far from amounting to a denial of loss, patently supports such a finding. The only question then is how we measure that loss. Given that the courts have traditionally been reluctant to stray from valuation

[51] *Wrotham Park Estate Co Ltd v Parkside Homes Ltd* [1974] 1 WLR 798 (Ch D).

[52] *Ibid*, at 815.

[53] See, too, R Sharpe and S Waddams, 'Damages for Lost Opportunity to Bargain' (1982) 2 *OJLS* 290; *Jaggard v Sawyer* [1995] 1 WLR 269 (CA); *cf Attorney-General v Blake*, above n 30, at 281 (Lord Nicholls).

[54] It was found that the claimant would not have consented to any such release.

[55] See, eg, A Burrows, *The Law of Restitution*, 2nd edn (London, Butterworths, 2002) 477; Edelman, above n 1, at 99–101.

determined objectively by reference to the market, it is entirely unsurprising that they have approached this question by asking what sum the claimant could reasonably have demanded in return for releasing the defendant from her obligation, since the price which a hypothetical claimant would accept in return for allowing the defendant to 'breach' identifies the value the claimant places on the defendant abiding by her obligation. To object that this is not the sum of money which the claimant would have charged is simply to object to adopting an objective, market-based assessment of loss, rather than one which takes account of the claimant's own priorities and preferences. As I have argued earlier,[56] in principle, loss should indeed be viewed, and hence identified and quantified, on a subjective basis. But this just means that our assessment of loss in these cases has been defective, not that a compensatory analysis must be regarded as fictitious and so to be rejected.

Once we understand that a loss is properly to be regarded as a reduction in the claimant's quality of life, we can see that the point of compensatory awards is to do something for the claimant which she values and so which improves her quality of life, such that, as a result of the award she ends up in as good a position as she would have occupied had the contract been performed. In this way, though the claimant has been deprived of one thing she values—performance of the contract—she gets something else which improves her quality of life by an equivalent degree. As I have been stressing over the previous paragraphs, loss should not be viewed as dependent on and co-extensive with a diminution in the assets at the claimant's disposal. As such, when it comes to making good losses, there is no reason for such an award to involve the payment of money or the transfer of some other asset to the claimant. In principle, providing the claimant with any 'thing' she values equally to performance (and any consequential losses) would suffice. Nonetheless, money clearly provides the most convenient means of compensation. Though individual's tastes and priorities differ, money is necessarily valuable to all of us, since it provides a ready means through which we can acquire those specific 'things' we want or need to pursue our objectives.

This then explains why we are right to leave claimants free to use the money they receive as compensatory damages as they wish. Unlike awards justified on the basis of the claimant's right to performance, the aim of compensatory awards is not to correct the specific harms caused by the defendant's breach but rather to ensure that the defendant does not leave the claimant worse off as a result of that breach, and hence as a result of suffering those specific harms.[57] A money payment does this because it

[56] See text accompanying n 47.

[57] It is because of this that suggestions that compensatory awards for 'non-pecuniary' losses are necessarily artificial or can only be based on convention are misguided: see, eg,

allows the claimant to purchase other assets or experiences which she values, and so which bring about an improvement to the claimant's position sufficient to offset the worsening of her position caused by the defendant. Accordingly, there is no sense in placing any limits on the claimant's use of the money she receives as compensation. Indeed, if the award is to have it intended effect, the claimant must be free to spend the money as she likes.

VI. SUBSTITUTIVE DAMAGES

Finally I want to say a little about an argument which is gaining a fair amount of support among commentators and which, if true, would provide a further justification for an award of substantial damages in the event of a breach of contract. This is that damages are and should be available as a substitute for the right infringed by the claimant, with such damages then being labelled 'substitutive' damages.[58]

The argument is that where a claimant's right has been infringed she should be entitled to damages from the defendant assessed by reference to the 'value' of that right, measured objectively and at the date of the breach.[59] Such awards do not depend on showing that any loss has been

Burrows, above n 44, at 29; *Wright v British Railways Board*, above n 49, at 777 (Lord Diplock). If we are not seeking to put right the specific physical harms suffered by the claimant, then all losses are fungible, since they all lie in the impact of the defendant's conduct on the claimant's overall quality of life. Though evidential difficulties may make it harder to estimate the gravity of this impact in some cases than in others, in principle all losses, whatever their source, can be assessed in monetary terms. This is not to say that claimants should always be entitled to recover the full sum of money it would take to bring about an improvement to their quality of life sufficient to offset the reduction caused by the defendant. Other considerations and principles may well justify restricting recovery to a lesser sum. In particular, where a claimant has been caused serious injury or illness, the impact on her quality of life will usually be so great that it can be offset only by the payment of a huge sum of money. Requiring a defendant to hand over such a sum would often be financially crippling and, at least where any blame to be attached to the defendant is minimal, may be considered unjust: see, eg, *Phillips v The South Western Railway Company* (1879) 4 QBD 406, 407 (Cockburn CJ). However, we should view these and similar considerations as going not to the identification or measurement of the claimant's loss but to the necessarily secondary question of how much of that loss the defendant should be required to bear.

[58] The language of 'substitution' and of 'substitutive' (or 'substitutionary') damages has also been used to describe other types of award. Some have used it to describe conventional compensatory awards (see the references above n 42). Smith, above n 10, identifies as 'substitutionary damages' awards which I have referred to here as 'performance interest' damages. It is important to distinguish these uses of the language of 'substitution' (and the awards they describe) from that discussed in the text.

[59] See, eg, R Stevens, *Torts and Rights* (Oxford, Oxford University Press, 2007) 59–84; M McInnes, 'Gain, Loss and the User Principle' (2006) 14 *Restitution Law Review* 76; *cf* S Elliott and C Mitchell, 'Remedies for Dishonest Assistance' (2004) 67 *MLR* 16; J Edelman, 'Gain-Based Damages and Compensation' in A Burrows and Lord Rodger (eds), *Mapping the Law: Essays in Honour of Peter Birks* (Oxford, Oxford University Press, 2006) 141, 153–8;

caused to the claimant, and indeed exist independently of any claim the claimant may also have for the recovery of such losses. Instead they follow simply from the infringement of the claimant's rights.

A number of points can be made. Firstly, we can ask why we should be seeking to a put a value on the claimant's rights. The language of 'value' is usually employed when we are looking to identify and assess losses and gains, as when we are concerned with awarding compensatory damages for losses caused by the defendant's breach. Yet such awards are heralded as noteworthy and distinctive precisely because they are different from, and exist independently of, standard compensatory claims. It is claimed that substitutive damages are awarded not to make good the consequences of the defendant's infringement of the claimant's right but simply as a reflection of the right that has been violated. Where then does value come into it?

One answer would be that substitutive awards *are* aimed at compensating a loss; it is just a different sort (or sense) of loss from that which is the focus of 'standard' compensatory awards.[60] So, rather than seeking to compensate the factual losses flowing from the defendant's breach, substitutive awards seek to make good the loss of the claimant's *legal rights* inherent in that breach. On this basis, we are looking to value the claimant's right precisely because this is the loss the claimant has suffered.

However, I doubt whether such a reformulation of the claimant's loss is successful. Firstly, we can question the premise that an infringement of a claimant's right entails or results in a loss of that right. Though breach can cause us to lose the factual advantages which rights seek to secure for us, I do not think it is quite right to say that we thereby lose the rights themselves. I think a better account is provided by Lord Clyde in *McAlpine v Panatown*:

> [W]hile frustration may destroy the rights altogether so that the contract is no longer enforceable, a failure in the obligation to perform does not destroy the asset. On the contrary it remains as the necessary legal basis for a remedy. A failure in performance of a contractual obligation does not entail a loss of the bargained-for contractual rights.[61]

So, where a duty is breached, and hence a right infringed, far from regarding that right as lost, the law's principal response should be to give

D Pearce and R Halson, 'Damages for Breach of Contract: Compensation, Restitution and Vindication' (2008) 28 *OJLS* 73. For further analysis and criticism of this argument: see A Burrows, 'Are "Damages on the *Wrotham Park* Basis" Compensatory, Restitutionary, or Neither?' in D Saidov and R Cunnington (eds), *Contract Damages: Domestic and International Perspectives* (Oxford, Hart Publishing, 2008) 165.

[60] See McInnes, *ibid*, at 84–6; *cf* Stevens, *ibid*, at 61–2, 78.

[61] *McAlpine v Panatown*, above n 30, at 534. See too Weinrib, above n 33, at 12–13 and Ripstein, above n 39, at 1971–72, 1978–82.

effect to that right, whether through specific enforcement or through a damages award such as that marked out above. And even where performance is no longer possible, I do not think that it follows that we should view the right as having been extinguished, as opposed, for instance, to concluding that it simply has no further bearing on the parties' future conduct.

In any case, even if we do consider that it is correct to say that the claimant's rights have been lost or damaged, there is then the question of how we value that loss. The law recognises us as having particular legal rights precisely because it considers that we should obtain certain factual advantages (and avoid certain factual harms or disadvantages). So we have a right to performance of contracts because the law wants to secure for us the factual advantages of such performance. But then the obvious measure of the claimant's lost right is by reference to the factual advantages of which she has been deprived by reason of the defendant's failure to respect that right. However, if this is true then the argument for substitutive damages does not justify awarding the claimant any more than the sum to which she is entitled straightforwardly as compensation for her factual losses. Conversely, if, as claimed, substitutive damages are to justify an award in excess of the claimant's factual losses, we need to be able to say why we should attribute a value to the claimant's right to performance in excess of the value of the factual advantages which that performance would secure. Simply to say that substitutive damages are assessed objectively and at the date of breach, whereas compensatory damages in respect of factual losses are to be assessed subjectively and at the date of judgment, and that this is why they can give us different figures, only prompts the question of why we *should* adopt different modes of assessment in each case.

A second version of the substitutive damages argument does not seek to explain such awards on the basis that they reflect any sort of loss to the claimant. But then we have no obvious answer to the question of why we should be concerned with putting a value on the claimant's right to performance. A court is confronted by a claimant who has suffered a breach of contract; her rights have clearly been infringed. But how do we get to the conclusion that the court should then award the claimant the value of that right? To say that damages should be ordered as a 'substitute' for this right or to put the claimant in the 'next best' position to actually receiving performance gets us no nearer to an answer.[62] In what way is such an award of damages a substitute for performance? Why should we consider this, rather than some other response, the next best way of dealing with breach?

[62] *Cf* Stevens, above n 59, at 59–60.

The fact that the claimant has a right to performance gives us a justification for an award of damages where and to the extent that this gives her the performance to which she is entitled. The claimant can claim that sum because that gives genuine effect to her right. She gets the very factual advantage—performance—that the right exists to secure for her. That cannot be said of a substitutive damages award of the kind described here. Nor can such an award be justified on the basis that it ensures that the claimant is not left worse off as a result of not receiving the performance to which she is entitled, since proponents of substitutive damages avowedly claim that such awards permit recovery in excess of the claimant's real, factual losses.

Since substitutive damages neither give claimants the performance which they are due and for which they have contracted, nor make good losses caused by the defendant's breach, we are left in search of any justification for them. The basic question is simple: why should claimants be entitled to recover a sum of money which is in excess of their losses and which in no way gives them the performance to which they are entitled? I am not intending to suggest that we cannot justify damages awards other than on the two bases I have put forward here.[63] My claim is simply that any such award must be supported on some basis other than that it vindicates or effectuates the claimant's right to performance.

One possible argument is that, beyond the question of protecting the claimant's rights, breach of contract threatens the very practice of contracting and the various broader moral or political values which contract law may be thought to support and embody. This may be what is meant by those who argue that we need to take performance, and the right to performance, of contracts seriously, or that an award of damages should reflect the seriousness of the right.[64] However, if we want an award or order that amounts to a clear statement that the claimant's rights have been infringed then this can be done by a declaration or an award of nominal damages. If we want to send out a stronger message that contracts are an important social and legal institution, and hence that contractual rights should be respected, then this may give us a further ground for an award of substantial damages. If so, the assessment of such an award should depend on a consideration of how this message may best be conveyed. There is no reason to think that this sum should bear any relation to the value which

[63] In particular, I have said nothing on the question of if and when we can justify damages awards assessed by reference to the gains made by the defendant as a result of her breach. The analysis presented here bears on this issue only to the extent that it tells us that such awards cannot be justified on the basis that they give effect to the claimant's right to performance. Whether some other justification for such awards can be found is a question I leave open.

[64] *Cf* Stevens, above n 59, at 78–9, 84, argues that the proper measure of substitutive damages turns upon the 'seriousness' of the infringement of the claimant's right.

may be attached either to the performance the claimant is due or to the right which aims to secure her that performance.

VII. CONCLUDING REMARKS

I think there is an important lesson here for those of us who are concerned with the justification of the awards and orders granted by courts not only in relation to claims following breaches of contract but in any aspect of private law. The recognition of legal rights follows from some understanding that the claimant deserves to be accorded certain factual advantages (or to be spared certain factual harms). The claims the law supports and the orders and awards courts may make then seek to ensure, and are justified to the extent that they do ensure, that the claimant then obtains the very factual advantages which her rights are designed to secure for her. If we are to argue that the law should recognise a certain type of claim or accommodate a certain type of award, we need to be able to say what right of the claimant such a claim or award effectuates. This requires us to show how such a claim or award is effective in bringing about the factual advantages which that right seeks to secure. If we are, in turn, to be able to defend a particular view of what rights the law should accord to individuals we need to be able to say why the law should be concerned with securing the relevant factual advantages for the identified class of right holders. Any such argument necessarily takes us into the fields of morality, economics, politics and the like—in essence, wherever we consider that the norms and ideals which shape and justify the law are to be found.

In this sense, this article has only touched the surface of the question of how the law should respond when faced with claims following breaches of contract. Much work needs to be done in explaining why a claimant should be accorded a right to performance, and indeed in examining whether this is a necessary consequence of all contracts. Nor have I said anything on the question of the principles which may challenge the recognition and protection of a right to performance, and hence which may limit the claims we make available to contract claimants.[65] For now, it is enough to say that recognition of a right to performance requires that we do not limit claimants to recovery of losses caused by the defendant's breach, but actually give them, where possible, the performance to which they are entitled. This can be done through specific enforcement of the contract. Alternatively, the right to performance can, sometimes, be effectuated through an award of damages which the claimant uses to

[65] Here we may also note that there are clearly principles, such as those underlying the rules on remoteness and mitigation, which limit the scope of the claimant's right to recover compensation for losses caused by the defendant's breach.

purchase an equivalent 'performance' from an alternative source. If we are to argue for awards beyond these, we must look elsewhere for our justification.

7

Damages and the Right to Performance: A Golden Victory *or Not?*

ROBERT STEVENS

I. INTRODUCTION

A STORY. When I was to be married, the task of purchasing the wedding rings fell to me. I purchased a pair of rings from a jeweller in London's Hatton Garden, a gold ring for me and a ring sold to me as made of platinum for my wife to be.

Over the years my wife's wedding ring changed colour. Instead of the steely-white of platinum it became yellowy gold. She took it to another jeweller who explained that her ring was not platinum but gold, plated in rhodium so that it looked like platinum. It was worth much less than the ring I had thought I was buying.

Far from being upset, my wife was pleased. She thought that the cheaper 'defective' wedding ring much more accurately represented the state of our relationship than any upmarket ring. She also, I suspect, likes it both as a talking point and as a reminder to me that I am not as smart as I like to think I am. Her happiness makes me happier.

In law, what remedy is available to me against the jeweller (who I shall assume made an honest mistake)? I cannot get my money back: my wife would never be parted from her wedding ring and would not now want the sort of ring I had intended. I am not in any way factually worse off as a result of the jeweller's breach of contract. The capital or re-sale value of the ring is lower than the one I was promised, but the ring is my wife's not mine, and was initially given by the jeweller to my best man to hold for her, not to me. In fact the ring as sold has left both me and my wife happier and factually better off than if I had had the ring I was promised. It has even proved unexpectedly useful in my job of teaching and explaining the law.

Even though I have suffered no actual loss I am still entitled to damages, assessed as the difference in value at the time of performance between the

ring I was promised and the ring delivered.[1] This article seeks to defend this result on the basis that damages are awarded, in the first instance, to vindicate the right violated, here the contractual right to a wedding ring of a particular quality. Such damages need to be differentiated from damages awarded to compensate for loss consequent upon the breach of a contract. The failure to differentiate between damages awarded as a substitute for the contractual right and those awarded to compensate for consequential losses suffered, leads to confusion and muddle.

II. WRONGS

A breach of contract is a wrong. A wrong is a breach of a duty. A breach of a duty is the infringement of a right. That contracts give rise to primary rights of performance can be demonstrated in a number of ways. For example, if a contract only gave rise to a right to performance *or* a right to damages for failure to perform, there would be no doctrine of frustration because it is never, or almost never, impossible to pay a sum of money by way of damages. The remedies which compel the performance of the primary contractual obligations (injunctions, specific performance, and by far and away the most common remedy of all, the action for the agreed sum) similarly disclose the existence of the primary right to performance. The one and only 'interest' which a contract gives rise to is the 'interest' or right to counter-performance.[2] The infringement of the primary right to performance gives rise to a secondary right to damages which did not exist prior to breach.[3] The existence of damages whose role is to vindicate the right to performance which has been infringed, and not to compensate for loss suffered as a result of the infringement, is one way of revealing the importance attached to the performance of what has been promised.[4]

Many other wrongs entitle the victim to claim substantial damages by way of substitution for the right infringed.[5] '[E]very violation of a right imports damage.'[6] Two examples will be given here.

[1] Sale of Goods Act 1979, s 53(3).

[2] D Friedmann, 'The Performance Interest in Contract Damages' (1995) 111 *LQR* 628; *contra* C Webb, 'Performance and Compensation: An Analysis of Contract Damages and Contractual Obligation' (2006) 26 *OJLS* 41.

[3] *Photo Production Ltd v Securicor Transport Ltd* [1980] AC 827 (HL).

[4] See also D Pearce and R Halson, 'Damages for Breach of Contract: Compensation, Restitution and Vindication' (2008) 28 *OJLS* 73.

[5] R Stevens, *Torts and Rights* (Oxford, Oxford University Press, 2007) 59–91. See also Lord Scott, 'Damages' [2007] *Lloyd's Maritime and Commercial Law Quarterly* 465.

[6] *Neville v London Express Newspaper Ltd* [1919] AC 368 (HL) 392 (Viscount Haldane). See also *Nicholls v Ely Beet Sugar Factory Ltd* [1936] 1 Ch 343 (CA) 350 (Lord Wright MR); *Jones Brothers (Hunstanton) Ltd v Stevens* [1955] 1 QB 275 (CA) 280 (Lord Goddard CJ).

The first, from 1703, is *Ashby v White*.[7] In that case, the claimant was prevented from voting by a constable on the pretext that he was not a settled inhabitant. The claimant's preferred candidate was elected and so he had no loss of any kind consequential upon the infringement of his right to vote, except perhaps some bruised feelings. In the Court of King's Bench a majority rejected the claim. Holt CJ dissented:

> [S]urely every injury imports a damage, though it does not cost the party one farthing, and it is impossible to prove the contrary; for a damage is not merely pecuniary, but an injury imports a damage, when a man is thereby hindered of his right.[8]

On appeal to the House of Lords, his dissent was upheld by a majority of lay peers. Harvey McGregor describes the approach of Holt CJ as 'fictional'.[9] However, without it, it is impossible to understand why the common law allows damages to be recovered without the proof of loss to the claimant or gain to the defendant.

The law has not settled on one label for this substitutive award. Sometimes the distinction between substitutive damages and consequential loss is described as that between 'general' and 'special' damages.[10] However, these terms are used in a variety of ways and are best avoided.[11] Again, sometimes the substitutive award is termed 'direct loss' and is contrasted with 'consequential loss'.[12] The label 'direct loss' is misleading, however, as the substitutive award is available even though no loss is suffered as a matter of fact. Substitutive damages are not compensatory for loss, properly so-called, at all.

Perhaps the most famous example in the law of torts is the decision of the House of Lords in *The Mediana*.[13] The plaintiffs operated a port. Their lightship was damaged as a result of the defendant's carelessness. The plaintiffs had a spare lightship to take the place of any ship damaged and

[7] *Ashby v White* (1703) 1 Bro PC 62, 1 ER 417 (HL), reversing (1703) 2 Ld Raym 938, 92 ER 126 (KB).

[8] *Ibid*, at 137.

[9] H McGregor, *McGregor on Damages*, 17th edn (London, Sweet & Maxwell, 2003) 360.

[10] *Ratcliffe v Evans* [1892] 2 QB 524 (CA) 528 (Bowen LJ).

[11] *Cf* J Jolowicz, 'The Changing Use of "Special Damage" and its Effect on the Law' [1960] *CLJ* 214; McGregor, above n 9, at 19–23.

[12] *Burdis v Livsey* [2003] QB 36 (CA) [87].

[13] *Mediana v Comet (The Mediana)* [1900] AC 113 (HL). See also *Steam Sand Pump Dredger v Greta Holme (The Greta Holme)* [1897] AC 596 (HL); *Mersey Docks and Harbour Board v Owners of SS Marpessa (The Marpessa)* [1907] AC 241 (HL); *Admiralty Commissioners v SS Chekiang (The Chekiang)* [1926] AC 637 (HL); *The West Wales* (1932) 43 Ll L Rep 504; *Edmund Hancock (1929) Ltd v 'Ernesto'* [1952] 1 Lloyd's Rep 467 (CA); *Lord Citrine v Hebridean Coast (The Hebridean Coast)* [1961] AC 545 (HL) [*The Hebridean Coast*]; *Birmingham Corporation v Sowsbery* [1970] RTR 84; *cf Alexander v Rolls-Royce Motor Cars Ltd* [1996] RTR 95 (CA); *Bella Casa Ltd v Vinestone Ltd* [2005] EWHC 2807 (TCC).

so were not forced to obtain a replacement whilst the damaged ship was repaired. There was no dispute that the defendants were liable to compensate the claimants for the cost of repairs, but were they additionally liable for the loss of the use of the lightship during the period of repairs? Although the loss of the use of the ship that had resulted had not left the claimant factually worse off, damages were payable representing the value of the right to use the ship during the period of repair. It is sometimes suggested that the principle in *The Mediana* should be confined to non-profit-making assets,[14] but the distinction between profit-making and non-profit-making assets was rejected by Lords Reid and Morton in *The Hebridean Coast*[15] and is inconsistent with the law as we find it.[16] Whether the right is wholly commercial is relevant, as we shall see, to its quantification, but this does not mean that commercial parties are confined to claims for consequential loss.

Why are damages awarded without proof of loss? The answer is that damages awards are the law's attempt to reach the 'next best' position to the wrong not having been committed. For breach of contract, this is the next best position to the performance having been rendered. The paper for the symposium at which this article was originally presented was due on 31 December 2007. By 1 January 2008 I could no longer keep my promise, but this did not relieve me from my obligation. If I can get a presentable article in by 2 January, that is what I must do. Damages are not owed in order to make good or eradicate losses, any more than is my obligation to make a late delivery coupled with an apology. Eradicating losses will frequently be impossible through a monetary award, as most obviously where the plaintiff has lost a limb or suffers terrible emotional distress as a result of the defendant's wrong. I cannot eradicate the stress of the organiser waiting for my article to arrive. Rather damages seek to achieve the closest position to the wrong not having occurred as can be ordered. The damages in substitution for the right which were awarded in *Ashby v White* or *The Mediana* represent the law seeking to achieve the next best thing to the right not having been violated in the first place. This may not be a particularly close approximation to the position which would otherwise have obtained, just as paying money for the loss of a limb will not cause it to grow back, but, given where we are, it is the best which can be ordered to be done.

[14] J Edelman, 'Gain Based Damages and Compensation' in A Burrows and Lord Roger (eds), *Mapping the Law: Essays in Memory of Peter Birks* (Oxford, Oxford University Press, 2006).

[15] Above n 13, at 577 (Lord Reid), 580 (Lord Morton); *contra The Bodlewell* [1907] 1 P 286; *cf Admiralty Commissioners v SS Valeria* [1922] 2 AC 242 (HL).

[16] See also *The Sanix Ace* [1987] 1 Lloyd's Rep 46 (QBD); *The Charlotte* [1908] P 206 (CA); and *The Winkfield* [1902] P 42 (CA).

Another way of demonstrating the importance of substitutive awards is where the defendant's action constitutes two separate wrongs to the plaintiff, which lead to different awards of damages. If an employer summarily fires an employee, the employee will have a claim for breach of contract. If, however, the breach also amounts to an act of discrimination, perhaps on the basis of sex or race, the plaintiff will have an alternative cause of action.[17] The loss suffered by the plaintiff may not differ according to whether the dismissal constitutes an act of discrimination or not. However, where it does so constitute discrimination greater damages are payable. The violation of this additional right leads to a higher award, despite the absence of any additional loss.

On one view, the mystery is not why the law awards damages based upon the value of the (contractual) right infringed, but why it goes further and allows the plaintiff to recover for his losses over and above the value of such right. Again, this is because it is the 'next best' to the wrong not having occurred. Paying money to make good consequential loss may not be the best that the defendant can do, but will usually be the best which can be ordered by a court to be done. For example, when I fail to do what I have promised, one aspect of the obligation to achieve the 'next best' is to apologise. The courts make money awards not because this is the next best which can necessarily be achieved but because this is the next best thing that they can order to be done. Involuntary apologies are not apologies at all.

All forms of damages awards for all wrongs, including breach of contract, seek to place the claimant in the position he or she would have been in if the wrong had not occurred, in so far as money can do this. There is no difference in this regard between damages for torts, equitable wrongs or breaches of contract. The underlying primary rights may be different in content, but the aim of the secondary obligation to pay damages for their infringement does not vary. This means that damages for breach of contract seek to place the claimant, so far as money can do it, in the position he or she would have been in if the contract had been performed. This is the next best position. In rare cases this may be incapable of proof. If, for example, an actor agrees to appear in a film, but in breach refuses to perform resulting in the film's cancellation, whether the film would have been a success may be impossible to demonstrate.[18] It cannot be shown what the position upon performance would have been. In such exceptional circumstances damages may be awarded which are based upon the position that the plaintiff would have been in if the contract had never been entered into. Such damages are the 'next-next-best' which can

[17] See Sex Discrimination Act 1975 (UK); Race Relations Act 1976 (UK).
[18] See, eg, *Anglia Television Ltd v Reed* [1972] 1 QB 60 (CA).

be awarded, and are only ordered where the next best position is unknowable.[19] However, for the purposes of this article the important point is that the 'next best' is not necessarily calculated by reference to loss suffered.

Not all wrongs give rise to an entitlement to substantial damages without proof of consequential loss. Deceit, for example, requires proof of consequential loss and the only remedy available is one to make good the loss suffered. Lies, per se, are not actionable. Slander, subject to exceptions and unlike libel, is similar. That breach of contract does not fall into the category of wrongs which require the proof of consequential loss before substantial damages will be awarded, is probably most readily demonstrated in the context of the sale of goods.

III. SALE OF GOODS

A. Non-delivery

Where a seller in breach of contract fails to deliver, the buyer is entitled to the difference between the market value of the goods and the contract price.[20] This rule cannot be displaced by demonstrating that the buyer suffered no loss consequent upon the seller's breach. This was established by the House of Lords in *Williams Bros v Ed T Agius Ltd*.[21] The defendant failed to deliver a consignment of Norwegian coal which it had agreed to sell to the plaintiff for 16s 3d per ton CIF Genoa (November shipment). The plaintiff had agreed to sub-sell coal of the same quantity and description for 19s to X. It was intended that the cargo the plaintiffs bought from the defendant would be used to fulfil the contract with X but they were under no obligation to do so. The market price at the time of breach stood at 23s 6d.

The defendants argued that where the plaintiff has in fact suffered less than the difference between the contract price and the market price at the time of the breach, then he is not entitled to the full market price. The plaintiffs had not bought in substitute goods at the higher market price and had no intention to do so, *and* there was no chance of their being sued for

[19] *C & P Haulage v Middleton* [1983] 1 WLR 1461 (CA).

[20] Sale of Goods Act 1979, s 53(3). The sale of goods legislation is a codification of the common law of contract in this context.

[21] [1914] AC 510 (HL) [*Williams Bros*]. See also *Diamond Cutting Works Federation Ltd v Triefus & Co Ltd* [1956] 1 Lloyd's Rep 216 (QBD) (Barry J); *Mouat v Betts Motors Ltd* [1959] AC 71 (PC NZ); *Derby Resources AG v Blue Corinth Marine Co Ltd (The Athenian Harmony)* [1998] 2 Lloyd's Rep 410 (QBD) (Colman J). But see *Sealace Shipping Co Ltd v Oceanvoice Ltd* [1991] 1 Lloyd's Rep 120 (CA) criticised by GH Treitel, 'Damages for Non-Delivery' (1991) 107 *LQR* 364.

non-delivery by the sub-buyer.[22] The House of Lords refused to discount the damages payable, applying the market value rule.

If, as is commonly thought, contract damages are only awarded to compensate for actual loss suffered, ignoring the sub-sale contract wrongly over-compensates the plaintiffs.[23] What *William Bros* demonstrates is that compensating for loss is not the sole purpose of contract damages.

Applying the market value rule, and ignoring sub-sales, might be thought to be justified by the goal of commercial certainty. The task of fixing the appropriate level of damages is made easier by ignoring the sub-sale. However, that this is not the true justification for the rule can be demonstrated by those cases where the sub-sale is taken into account because the *consequential loss* the plaintiff suffers as a result of the terms of the sub-sale is higher than would normally have been suffered. Normally, in the case of generic goods for which there is a ready market there will be no consequential loss, as the buyer will be able to go into the market and fulfil the sub-sale contract.[24] However, this will not always be so. This may be illustrated by another decision of the House of Lords, *Re Hall (R and H) Ltd and Pim (WH)(Jnr) and Co's Arbitration*,[25] whose facts closely resemble *Williams Bros*. On 3 November, the defendant agreed to sell to the plaintiff a cargo of Australian wheat CIF UK ports at 51s 9d a quarter. On 21 November, the plaintiffs sub-sold wheat of the same quantity and description to X at 56s 9d. X, in turn, had agreed to sell such a cargo to Y at 59s 3d a quarter. The defendant bought a cargo of wheat on board *SS Indianic* at 60s a quarter. The defendant secured agreement with all concerned that the sales should each be treated as a re-sale of the cargo which was the subject matter of the preceding purchase in the chain. The defendant gave notice appropriating the *Indianic* cargo to its contract with the plaintiff and that notice was passed down the chain. The defendant sold the *Indianic* cargo to a third party at 59s 1/2d a quarter. When the cargo arrived, the market price was 53s 9d a quarter. Having sold the cargo, the defendant was unable to deliver the documents covering the cargo to the plaintiff. The Court of Appeal held that Hall's damages were limited to the difference between the market price (53s 9d)

[22] This was either because the terms of the sub-sale contract relieved the plaintiffs of their obligation to deliver when no delivery to them took place, or because X had in turn sold goods of the same quantity and description back to the defendant, assigning to the defendant its rights against the plaintiff under the contract of sub-sale. In order to rely upon the rights under the assigned sub-sale the defendant should have raised a counter-claim.

[23] AS Burrows, *Remedies for Torts and Breach of Contract*, 3rd edn (Oxford, Oxford University Press, 2004) 213; *cf* SM Waddams, *The Law of Damages*, 3rd edn (Aurora, Canada Law Book, 1997) 1.1360–1.1440, 1.1930, 1.1940.

[24] See *Kwei Tek Chao v British Traders & Shippers Ltd* [1954] 2 QB 459, 489–90 (Devlin J).

[25] (1928) 33 Com Cas 324 (HL). See also *Household Machines Ltd v Cosmos Exporters Ltd* [1947] KB 217.

and the contract price (51s 9d) at the date of the breach. The plaintiff claimed the difference between the price under their sub-sale to X (56s 9d) and the contract price (51s 9d).

The House of Lords restored the decision of Rowlatt J that the plaintiff was entitled to recover the difference between the price at which it had bought and the price at which it had resold the cargo together with an indemnity for the damages and costs which the plaintiff would have to pay to the buyers who had bought from them. Where what is claimed is consequential loss over and above the market value of the goods this should be recoverable unless it is too remote. It was well known to both parties that it was common practice to resell cargoes whilst afloat. Apart from common knowledge, the contract itself showed this. Moreover, the correspondence as to the actual appropriation of the vessel was additional proof, if proof were needed, of the familiarity of the defendant with the practice of successive re-sales of cargo afloat. The defendant knew as soon as it nominated a cargo that only delivery of that cargo could satisfy the contract, and it was sufficient to give rise to liability for loss of profit that there was an even chance of a sub-sale taking place.

It is not correct therefore that sub-sales are ignored. If the sub-sale increases the loss suffered as a result of the defendant's breach, it should be recoverable unless too remote.[26] However, where the actual loss suffered is lower than the market value because of the sub-sale, this does not reduce the damages payable as the buyer can claim damages based upon his or her right to performance regardless of the consequential loss actually suffered.

The justification for the market value rule in *Williams Bros* is not, therefore, the promotion of a commercially certain rule fixing damages at a particular point. Sometimes sub-sales are ignored and sometimes they are taken into account. Further, the rule is not justified on the basis that 'what is sauce for the selling goose is sauce for the buying gander', as is also sometimes suggested.[27] The fact that sub-sales can be brought into account by the buyer to increase the damages payable over and above the market value assessment, but cannot be used by the seller by way of reduction, is only explicable on the basis that sometimes damages are measured by the loss suffered, and sometimes they are not.

The Ontario Law Reform Commission proposed to overturn the common law in cases of non-delivery and confine the claimant to damages for actual loss suffered, reversing *Williams Bros*.[28] If piecemeal reform of this

[26] For cases where such loss was too remote: see *Williams v Reynolds* (1856) 6 B & S 495, 122 ER 1278; *Aryeh v Lawrence Kosteris & Son Ltd* [1967] 1 Lloyd's Rep 63 (CA); *Burgoyne v Murphy* [1951] 2 DLR 556 (NBSC AD).

[27] *Williams Bros*, above n 21, at 523 (Lord Dunedin).

[28] Ontario Law Reform Commission, *Report on Sale of Goods* (Toronto, Ministry of the Attorney General, 1979) 502.

kind is adopted it will create anomalous differences between damages in the law of sale and in other areas, such as carriage.

B. Defective Goods

Until recently, the position in England in relation to defective goods could be confidently stated to be consistent with the law on non-delivery, as illustrated by the leading case of *Slater v Hoyle & Smith Ltd.*[29] In that case, the plaintiff buyers bought from the defendant sellers a quantity of unbleached cotton cloth, some of which they then used, after bleaching, to fulfil a contract of sale to a third party. The cloth supplied was of inferior quality to that warranted under the contract. The price under the sub-sale contract was paid in full, and no claim was brought by the sub-buyers against the plaintiff. The sellers argued that the buyers had suffered no loss as a result of the defect in quality. The Court of Appeal upheld the award of damages based upon the difference in market value at the time of delivery between the defective cloth and cloth conforming to the quality requirements of the contract.

Various arguments can be constructed to attempt to show that the buyers did in fact suffer a loss, but none is convincing. For example, it could be said that the buyers were not bound to deliver the same goods under the sub-sale that they received under the contract of sale. The buyer could have purchased other goods on the market to fulfil the sub-sale. However, in quantifying loss we are interested in what actually did happen: on the facts the buyers did not go out into the market in this way.

Similarly, it might be said that the buyers have paid too much for the goods, and that this loss is a real loss. However, damages are not quantified by reference to what, if anything, the plaintiff overpaid at the time of contracting but rather by reference to the value of the contractual right at the time of performance. If, for example, £100 is paid for cloth which, as delivered, has a value at the time of contracting of £80 or which the plaintiff would have been prepared to pay £80 for if he or she had known of its quality at time of contracting, damages are not calculated by looking at the difference between what was paid and what would have been paid. If, therefore, the market rises so that at the time of delivery the market value of the goods as promised is £120, but the value of the goods delivered is at that time £90, damages of £30 are payable, not £20.

Unfortunately, *Slater* was not followed by the Court of Appeal in *Bence Graphics International Ltd v Fasson UK Ltd.*[30] The defendants sold a quantity of vinyl film to the buyer who used it to make decals that were

[29] [1920] 2 KB 11 (CA) [*Slater*].
[30] [1998] QB 87 (CA) [*Bence*].

used in the shipping industry to identify bulk containers. It was a requirement of the container trade that the containers would last for five years, so it was stipulated in the contract of sale that the film would last for five years, so that what was printed upon it would last for the industry standard period of time. However, the sellers supplied sub-standard vinyl which did not survive the contracted-for period. The buyers received extensive complaints from end-users. This resulted in only one minor claim that the buyers settled and for which they were reimbursed by the sellers. The trial judge awarded damages by reference to the market value of what was delivered. The Court of Appeal by a majority allowed the appeal, awarding a small sum with respect to that film which was unused and remitting the case for the assessment of consequential loss, based upon the plaintiff buyer's liability to the ultimate end-users of the film.

The case has been convincingly criticised,[31] and even its supporters concede that the reasoning of the majority is unconvincing. First, the majority treated the question in issue as one of remoteness. This was a mistake. Where damages are claimed for the difference in value between what was promised and what was delivered, no issue of remoteness arises.[32] Only where the claim is for *consequential loss* does the issue of remoteness arise. Second, the majority failed to provide any convincing distinction between *Bence* and *Slater,* the latter being a unanimous decision of Bankes, Warrington and Scrutton LJJ. Otton LJ sought to distinguish *Bence* on two narrow bases: first, that the goods sold on in *Slater* were substantially the same, whereas in *Bence* they had been substantially processed; and second, that in *Slater* the seller had not known of the contemplated re-sale whilst in *Bence* they had. The first ground is both unconvincing on the facts, and of no relevance to the basis of the market rule. The second is also questionable as a matter of fact, difficult to normatively understand, and even if accepted, would indicate exactly the opposite results: substantial damages in *Bence* and no recovery in *Slater*. Auld LJ was bolder still, failing to distinguish *Slater* but refusing to follow it.

Bence is wrong in reasoning and result. It is flatly inconsistent with prior Court of Appeal authority and inconsistent with the rationale behind the rule in cases of non-delivery established by the House of Lords in *Williams Bros*. In Canada *Bence* can be safely ignored as, on facts materially identical to *Slater,* the Supreme Court in *Bainton v John Hallam Ltd* reached the same result as had the English Court of Appeal five months earlier.[33]

[31] GH Treitel, 'Damages for Breach of Warranty of Quality' (1997) 113 *LQR* 188; C Hawes, 'Damages for Defective Goods' (2005) 121 *LQR* 389; MG Bridge, 'Defective Goods and Sub-sales' [1998] *Journal of Business Law* 259.

[32] Treitel, *ibid*, at 190.

[33] [1920] 60 SCR 325, 54 DLR 537. *Slater* was not cited, presumably because it was not yet reported.

Superficially difficult to reconcile with *Slater* is the earlier decision of the House of Lords in *British Westinghouse Electric and Manufacturing Co Ltd v Underground Electric Railways Co of London Ltd*.[34] In 1902 the defendants agreed to sell eight steam turbines for electricity generation to the plaintiff railway company for £250,000. The turbines proved to be defective. The plaintiff sought damages of upwards of £280,000 for losses that they estimated would be caused by the excessive coal consumption over the life of the machines. Alternatively, they claimed the cost of installing eight new turbines, with superior kilowatt capacity, which they had purchased in 1908 when the seller's machines proved insufficient. They estimated the cost of these machines to be £78,186. Even if the seller's machines had complied with the conditions of the contract, it would still have been to the advantage of the buyers (at their own cost) to have replaced the machines supplied with the new machines as soon as the latter became available on the market.

It should be apparent that the claim for damages as framed was one for *consequential loss*. These losses were not in fact suffered because 'the superiority of the [replacement] machines and of their efficiency in reducing working expenses was in point of fact such that all loss was extinguished, and that actually the respondents made a profit by the course they took'.[35] The claim was not one for the difference in market value between the machines promised and those delivered. It was not a claim based upon the value of the right to performance. Indeed such a claim could not have been brought in addition to that which was asserted. The plaintiff had recovered substantial damages for the losses incurred because the machines were defective during the years up until their replacement. A plaintiff cannot recover *both* the difference in value between what it was promised and what it received, *and* the expense it in fact incurs in making good the defective performance. Recovering the former means that the latter loss is, to that extent, not incurred. If my wife had insisted upon a new platinum ring I could not have claimed by way of damages both the difference in value between the ring promised and delivered, and in addition the cost of replacing it. If I recover the former, to that extent, I do not suffer the latter. Recovery under one head reduces the damages recoverable under the other.

Where, however, a claim for damages based upon the value of the contractual right is brought, the issue of mitigation which arose in *British Westinghouse* does not arise. Questions of mitigation, like remoteness, only arise in relation to consequential losses. So, *British Westinghouse* may

[34] [1912] AC 673 (HL) [*British Westinghouse*]. See also *Erie County Natural Gas & Fuel Co v Carroll* [1911] AC 105 (PC Can); *Cockburn v Trusts & Guarantee Co* (1917) 55 SCR 264, 37 DLR 701.

[35] *British Westinghouse*, *ibid*, at 688.

be usefully contrasted with the subsequent decisions of the Privy Council in *Jamal v Moolla Dawood Sons & Co*[36] and the Court of Appeal in *Campbell Mostyn (Provisions) Ltd v Barnett Trading Co*.[37] Consequential losses are assessed at the time of judgment, taking into account events which have happened subsequent to the wrong that increase or decrease the loss which would have been expected to have been suffered. This was the approach in *British Westinghouse*. By contrast, in assessing damages which are substitutive for the right, the valuation of the right takes place at the time of infringement.[38] In cases of non-delivery, or the delivery of defective goods, this is at the time of delivery. Subsequent events are irrelevant as the court's task is not to calculate what actual loss has been suffered. The correct question is: what is the difference in value at the time of breach between the performance promised and the performance received?

In *Jamal* the contract concerned the sale of a quantity of shares for 185,000 rupees, with delivery to take place some months later on 30 December 1911. The shares fell substantially in value and the defendant buyer refused to complete. Subsequent to the buyer's repudiation, the seller began to sell the shares elsewhere. The seller managed to sell at prices significantly higher than the market price at the time of breach, thereby avoiding the loss he would otherwise have suffered. The defendant buyer argued that this should be taken into account in quantifying damages. The Privy Council correctly rejected this contention, assessing damages by reference to the value of the right to performance at the time of breach.

Precisely the same is the decision of the Court of Appeal in *Campbell Mostyn*. The defendant buyer wrongfully rejected 350 tins of South African ham which they had agreed to purchase. Two weeks after the breach, the government announced restrictions on ham imports from Continental Europe. This action increased the price at which the plaintiffs were able to re-sell and resulted in the sellers suffering no loss as a result of the buyer's breach. According to the court, the plaintiff's successful mitigation of any loss they might have suffered did not reduce the damages payable, these being assessed by reference to the market price at the time of breach. Again, the actual loss suffered at the time of judgment was ignored because the damages here, unlike in *British Westinghouse*, did not seek to compensate for the actual loss suffered.[39]

[36] [1916] 1 AC 175 (PC) [*Jamal*]. Viscount Haldane LC who delivered the speech for the court in *British Westinghouse* was also a member of the panel in *Jamal*. For a sale of goods example: see *Jones v Just* (1868) LR 3 QB 197.

[37] [1954] 1 Lloyd's Rep 65 (CA) [*Campbell Mostyn*].

[38] See generally Stevens, above n 5, at 69–70.

[39] An unsatisfactory exception is made when, after a buyer has wrongfully rejected goods under a contract, and a new agreement is entered into by the same parties for the sale and purchase of the same goods at a reduced price. Where the buyer sues the seller for damages

C. Late Delivery

In *Slater*, Scrutton LJ was strongly critical of the earlier decision of the Privy Council in *Wertheim v Chicoutimi Pulp Co*.[40] However on the basis of the thesis put forward here both are readily reconcilable. The contract was for the sale of 3,000 tons of Canadian moist pulpwood FOB Chicoutimi for delivery by November 1900. The buyers had already contracted to sell 2,000 tons of pulpwood, and subsequently made contracts which accounted for the balance of the FOB quantity. The pulpwood was not in fact shipped until June 1901. The buyers were able to perform the sub-sale contracts despite the delay.

The market price prevailing had the goods been delivered on time was 70s, but it had fallen back to 42s 6d by the time delivery was actually made. The plaintiff sought to recover the difference in market value: 27s 6d. The sub-sale price of the goods was 65s and the plaintiff was able to perform these contracts despite the delay. In light of these facts, the House of Lords refused to make an award for loss which was not actually suffered.

What is the distinction between the late delivery in *Wertheim* and the non-delivery in *Williams Bros*? The diligent reader of the speeches of the House of Lords and the advice of the Privy Council will not find a satisfactory explanation. Again, the difference can be explained by recognising that damages may be awarded as a substitute for the right infringed. What was the difference in value between what was promised and what was received *at the time of breach*? More precisely, what was the difference in value between a delivery on 1 November and a June delivery for a person without knowledge of the future events, at the start of November 1900? At that time the market fall had not yet occurred, and so the difference would either have been nil or very small. If the plaintiff wished to claim damages for the loss suffered as a result of the market fall subsequent to the breach, this loss would actually have had to have been suffered which it was not.

One mystery does remain, however. The Privy Council did uphold an award of damages based upon the difference between the market price prevailing at the due date of delivery (70s) and the sub-sale price (65s). How this award of 5s was arrived at is a mystery as it neither reflects the value of the right nor the loss suffered. The best explanation may be that the plaintiff was entitled to some damages because of the violation of his

under the original contract, account can be taken of the profit made by the buyer on the subsequent contract, provided that it is found as a fact that the subsequent contract is part of a continuous dealing between the parties: see *Pagnan & Fratelli v Corbisa Industrial Agropacuaria Ltda* [1970] 1 WLR 1306 (CA).

[40] [1911] AC 301 (PC) [*Wertheim*].

right to timely delivery, and that the defendant had conceded that this sum was an appropriate level of damages.

IV. CARRIAGE OF GOODS

The position in relation to the sale of goods is also found in relation to their carriage. *Rodocanachi Sons & Co v Milburn Bros* concerned an action for non-delivery under a voyage charter of a cargo of cotton shipped at Alexandria on account of the charterer and bound for the United Kingdom.[41] Owing to the Master's negligence the cargo was lost. The charterer had sold on the cargo on a 'to arrive' basis, relieving the charterer of any obligation should the goods fail to arrive. The sub-sale was at a price considerably below the market price when the goods should have arrived in the United Kingdom. The contention of the defendant—namely, that the plaintiff ought not to be placed in a better financial position than would have obtained if the contract had not been broken—was rejected by the Court of Appeal. The value of the goods at the time and place of delivery was recoverable 'independently of any circumstances peculiar to the plaintiff, and so independently of any contract made by him for the sale of the goods'.[42]

Similar is the important decision of the House of Lords in *Ströms Bruks Aktie Bolag v Hutchison*.[43] The claim was for non-delivery under a voyage charter from Sweden to Cardiff of a cargo of wood pulp to be lifted in two shipments, the second in August–September at the defendant carrier's option. The contract was made to enable the plaintiff to fulfil a contract of sale which provided for the *delivery* of the cargo in Cardiff. The terms of the contract of sale and the contract of carriage did not therefore align, the plaintiff charterer's mistake appearing to have arisen from the fact that although the contract was described as being a CIF sale, it was by its terms one for the delivery of the goods. The defendant argued that even if they had performed their contract to the letter (that is, if the cargo had been lifted at the end of September) this would not have enabled the plaintiff to perform the contract of sale, the benefit of which would have been lost in any event. They should not, therefore, be liable for loss which would have been suffered anyway. The House of Lords rejected this argument. There was no need to prove 'special' damage. The contract of sale was simply the 'best evidence possible' of the general damages suffered by the plaintiff.[44]

[41] (1887) 18 QBD 67 (CA) [*Rodocanachi Sons*].
[42] *Ibid*, at 77 (Lord Esher MR).
[43] [1905] AC 515 (HL).
[44] *Ibid*, at 526 (Lord Macnaghten). See also *Biggin v Permanite* [1951] 1 KB 422, 438 (Devlin J); *Eldon Weiss Home Construction Ltd v Clark* (1982) 39 OR (2d) 129 (Co Ct) 133.

Put another way, it was the best available evidence of the value of the right to have the goods delivered in Cardiff at that date.

The decision of the Court of Appeal in *White Arrow Express Ltd v Lamey's Distribution Ltd* concerns the provision of a completed, if defective, carriage.[45] The Court of Appeal concluded that where the claimant has bargained for a deluxe delivery service and has received only a standard service, he or she is entitled to the difference in market value between what was received and what was bargained for, although the claimant failed to establish any such difference on the facts. Sir Thomas Bingham MR stated:

> It is … obvious that in the ordinary way a party who contracts and pays for a superior service or superior goods and receives a substantially inferior service or inferior goods has suffered loss. If A hires and pays in advance for a 4-door saloon at £200 per day and receives delivery of a 2-door saloon available for £100 per day, he has suffered loss. If B orders and pays in advance for a 5-course meal costing £50 and is served a 3-course meal costing £30, he has suffered loss. If C agrees and pays in advance to be taught the violin by a world famous celebrity at £500 per hour, and is in the event taught by a musical nonentity whose charging rate is £25 per hour, he has suffered loss.[46]

Whilst Sir Thomas Bingham MR describes the recipient of the service as having suffered a loss, this is not a consequential loss. In the examples given, the claimant is not, or at least is not necessarily, left in a factually worse position after performance than if he or she had received the bargained-for service. The 'loss' constitutes failing to receive performance of the bargained-for quality. His approach is correct and ought to be applied to all contracts for the provision of services just as it does to contracts for the sale of goods.[47] So if E contracts with F for the provision of an ordinary service, and receives an inferior one, E should be entitled to a claim for damages, just as should the person who has contracted for a superior version.

A Canadian decision which is apparently out of line with these principles is the decision of the Federal Court of Appeal in *Redpath Industries Ltd v The Cisco*.[48] The defendant carrier carelessly damaged the plaintiff charterer's cargo of sugar by seawater. The sugar as damaged was only sellable as animal feed for $53,000, whilst undamaged it would have a value of $280,000. The plaintiff was a sugar refiner who was able to refine small amounts of wet sugar with large amounts of dry so that it was sellable as

[45] (1996) 15 Tr LR 69.

[46] *Ibid*, at 73.

[47] See also *Attorney General v Blake* [2001] 1 AC 268 (HL) 286 (Lord Nicholls) [*A-G v Blake*].

[48] (1994) 110 DLR (4th) 583 (CA) leave to appeal to the SCC refused 116 DLR (4th) vii.

undamaged. This refining cost about $50,000. Damages were confined to the lower sum, the court treating the claim as one for a loss which had in fact been mitigated.

V. SALE OF LAND

Unsurprisingly, the position found in relation to contracts for the sale of goods is reflected in claims for breach in relation to contracts for the sale of land.[49] However, as land is not generic, specific performance is commonly available, leading to possible differences in relation to the timing of assessment of the value of the right.

Where specific performance or an injunction is awarded the court is not ordering a remedy for breach of contract but rather is compelling the performance of the primary contractual right. If, for example, a seller of land repudiates the contract, the buyer may reject his or her repudiation and seek performance. A repudiation which is not accepted is a thing writ in water,[50] which does not constitute a breach. Specific performance may be awarded even though there is no actual breach of contract.

Where damages are awarded 'in lieu of specific performance' or 'in lieu of an injunction' such damages are also substitutive. They are awarded in lieu of the right to performance. Again, proof of consequential loss is unnecessary.

This may be illustrated by the decision of the Supreme Court of Canada in *Semelhago v Paramadevan*.[51] The plaintiff agreed to buy a house from the defendant in the Toronto area for $205,000. The defendant seller reneged on the deal, and conveyed the property to a third party. The value of the property rose between the time of repudiation and trial to $325,000. However, the plaintiff kept his own house which rose in value during the same period from $190,000 to $300,000. The defendant argued that in calculating damages, the court should take into account the fact that had the contract been performed the plaintiff would not have acquired the benefit of the increase in value of his old property because he would have had to have sold it.

When damages are awarded in lieu of specific performance the correct moment for assessing the value of the right is the time of judgment. It is the right to performance at that time for which damages are awarded as substitute. This may be contrasted with a claim based upon a pre-trial

[49] See *Brading v McNeill & Co Ltd* [1946] Ch 145.

[50] See *Howard v Pickford Tool Co Ltd* [1951] 1 KB 417 (CA) 421 (Asquith LJ).

[51] [1996] 2 SCR 415, 136 DLR (4th) 1. See L Smith, 'Understanding Specific Performance' in N Cohen and E McKendrick (eds), *Comparative Remedies for Breach of Contract* (Oxford, Hart Publishing, 2005) 221, 228–30.

breach, where the right is valued at the time of breach.[52] The award in *Semelhago* was not assessed according to the gain the defendant made from the breach—an issue which the court did not examine. Indeed the gain made by the increase in value was made by the third party, not the defendant vendor. Rather the award monetarises the right to performance, even though this leads to an award over and above any loss suffered by the plaintiff.

It may be objected that awarding damages calculated by reference to the value of performance at time of trial is unfair to defendants as plaintiffs are allowed to speculate, without risk, on market movements. If the house goes up in value after the date upon which performance was due, plaintiffs can bring a claim and capture this gain for themselves, if the market declines they can leave the house with the defendants. However, there is no injustice to defendants once it is accepted that damages in lieu of specific performance are an indulgence: they are afforded where they do not wish to perform. If defendants wish to forestall speculation on the market they are always able to do so by performing their promises. Even as late as the date of trial, defendants can opt to perform rather than pay damages.

Lord Cairns' Act allows damages assessed according to the value of the right at the time of judgment to be awarded.[53] It is this ability which it adds to the court's armoury. It is unclear what purpose those who consider damages to be always compensatory think Lord Cairns' Act is serving.

VI. COVENANTS TO REPAIR

In *Joyner v Weeks* a tenant was in breach of a covenant of repair.[54] The landlord suffered no loss because he had entered into another lease under which the subsequent tenant agreed to affect repairs. If damages were only awarded for consequential loss, the landlord's claim ought to have failed as he had mitigated any loss after the tenant's breach. The claim for substantial damages succeeded. As the Court of Appeal stated many 'cases may be put in which it is plainly immaterial that at the commencement of an action for a breach of contract the plaintiff is in fact no worse off than he would have been if the contract had been performed'.[55] The value of the landlord's right to the work was quantified according to the difference in value between the repaired and unrepaired premises, capped by the cost of doing the repairs.

[52] See *Wroth v Tyler* [1974] Ch 30 (Megarry J) but see *Johnson v Agnew* [1980] AC 367 (HL).

[53] Chancery Amendment Act 1853, s 2 (now contained in Supreme Court Act 1981, s 50).

[54] [1891] 2 QB 31 (CA); *Smiley v Townshend* [1950] 2 KB 311 (CA); *Haviland v Long* [1952] 2 QB 80 (CA).

[55] *Joyner v Weeks*, *ibid*, at 33.

VII. BUILDING WORK

In *Alfred McAlpine Construction Ltd v Panatown Ltd*, Lord Goff gave the following example illustrating the same principle in the context of building work:

> [A] wealthy man who lives in a village decides to carry out at his own expense major repairs to, or renovation or even reconstruction of, the village hall, and himself enters into a contract with a local builder to carry out the work to the existing building which belongs to another, for example to trustees, or to the parish council. Nobody in such circumstances would imagine that there could be any legal obstacle in the way of the charitable donor enforcing the contract against the builder by recovering damages from him if he failed to perform his obligations under the building contract, for example because his work failed to comply with the contract specification.[56]

As Lord Goff makes clear the recovery of substantial damages is not dependent upon the plaintiff incurring any actual loss: 'Is it really to be suggested that his action will fail, because he does not own the hall, and because he has not incurred the expense of himself employing another builder to do the remedial work?'[57]

That Lord Goff's approach is correct is apparent from the analysis given in the previous sections of this article. However, two members of the court, Lord Jauncey and Lord Clyde, rejected his approach and *McAlpine v Panatown* has left the law in a state of uncertainty.

The facts of the case were as follows. Panatown employed McAlpine to build an office block and multi-storey car park on a site in Cambridge. Panatown alleged that the building was seriously defective. The site was not, however, owned by Panatown but by its holding company Unex Investment Properties Ltd ('UIPL'). On the same day that the construction contract was entered into, McAlpine signed a duty of care deed in favour of UIPL undertaking to take reasonable care and skill in carrying out the building contract.

Panatown sought to recover damages against McAlpine on two grounds. First, they argued that they were entitled to recover damages on behalf of the third party (the 'narrow ground'). This ground need not be considered further. Second, they argued that they were entitled to damages in their own right because they had not received what they were promised ('the broad ground'). The broad ground was accepted by Lord Goff and Lord Millett, who dissented, but rejected by Lord Jauncey and Lord Clyde. Lord Browne-Wilkinson who had expressed sympathy for the broad ground in *Linden Gardens Trust Ltd v Lenesta Sludge Disposals Ltd*[58] and conceded

[56] [2001] 1 AC 518 (HL) 547 [*McAlpine v Panatown*].
[57] *Ibid*, at 548.
[58] [1994] 1 AC 85 (HL) 112.

that examination by academic writers had discovered no serious difficulties with it, was not prepared to allow it to succeed where the third party had a right of enforcement in his or her own right.

In the example given by Lord Goff, it is submitted that there would, on any view, be a loss if, after the work had been done, the philanthropist reasonably employed a second builder to remedy the defective work or at least was likely to do so.[59] Can the promisee recover substantial damages even where this is not so? Lord Goff and Lord Millett would allow recovery in such circumstances. In the example given by Lord Goff, the third party would have no contractual right against the promisor, and so Lord Browne-Wilkinson may well also have been prepared to allow recovery.[60] Whether Lord Browne-Wilkinson was right to refuse to allow a claim for substantial damages in *McAlpine v Panatown* itself depends upon the correct construction of the agreement. Was what was bargained for an office block and multi-storey car park, or was what was bargained for such a building *or* a claim for damages by UIPL? Did the right to claim under the deed exclude the ability to claim damages on any other basis, just as would a liquidated damages clause? If the latter construction is adopted, the refusal of Panatown's claim for damages seems unobjectionable. However, on the more plausible construction of the parties' deal, the duty of care deed was required not in order to give UIPL a claim for damages in its own right, but rather for the benefit of subsequent purchasers of the premises.[61] On the better construction of the bargain, the claim for damages in *McAlpine v Panatown* ought to have succeeded.

Although at certain points Lord Goff (and Lord Millett) seem to indicate that the appropriate method for quantifying the claim for damages is the cost of repairs, this would be a mistake. Repair costs are a measure of loss. Where not actually incurred, such costs are only of relevance in valuing the right to performance as a matter of *evidence*. That damages are not awarded for actual expenses incurred is apparent from Lord Goff's treatment of damages for delay. The employer has 'a contractual right to the performance by the contractor of his obligation as to time, as much as he has to his performance of the work to the contractual specification'.[62] This right can be

> measured objectively in financial terms with reference to the anticipated profitability of the development; and this can provide an appropriate yardstick for

[59] *McAlpine v Panatown*, above n 56, at 533 (Lord Clyde), 574 (Lord Jauncey); *Radford v De Froberville* [1977] 1 WLR 1262 (Ch D).

[60] *Cf Woodar Investment Development Ltd v Wimpey Construction UK Ltd* [1980] 1 WLR 277 (HL) 300 (Lord Scarman); *Bovis Lend Lease Ltd v RD Fire Protection Ltd* (unreported, QBD, 6 February 2003) (Thornton J, QC).

[61] I Wallace QC, 'Third Party Damage: No Legal Black Hole?' (1999) 115 *LQR* 394; *McAlpine v Panatown*, above n 56, at 593 (Lord Millett).

[62] *McAlpine v Panatown*, *ibid*, at 554–5.

> measuring the estimated damages for delay in the performance for which the employer has contracted, even where the development was to be carried out on a site belonging to another person.[63]

Furthermore, to award the full difference in value between the building as promised and as built, may be incorrect. The plaintiff has a right to the work; the increase in value of the building by the work's completion may be greater than this because of the site's potential for improvement if the work is completed. Damages should therefore be capped at the difference in value between the work promised and the work performed.

The limits of the analysis presented here may be illustrated by considering the case of the disappointed legatees. In *White v Jones* a testator changed his will after a quarrel with his daughters so as to leave them nothing.[64] After reconciliation, he instructed his solicitor to draw up a new will leaving them £9,000 each. The solicitor negligently failed to change the will, resulting in the daughters inheriting nothing when he died. Although the estate of the testator should have a claim for substantial damages (since the bargained-for services had not been provided), this claim would not be calculated by reference to the sum that the legatees had failed to inherit—a sum which may be vastly greater than the value of the service bargained for. If a claim to the benefit that has not been conferred upon the legatees is to be allowed, it must be done on a different basis, either by allowing the estate to claim this sum on their behalf, or by allowing the legatees a claim in their own right as an exception to the doctrine of privity of contract.[65]

McAlpine v Panatown may also be usefully contrasted with the earlier decision of the House of Lords in *Ruxley Electronics and Construction Ltd v Forsyth*.[66] In that case, the plaintiff contracted for the construction of a swimming pool with a depth of 7 feet 6 inches. This was deeper than the industry standard because the plaintiff was a tall man who did not like to see the bottom of the pool when he dived in. The pool, however, was built to a depth of 6 feet 9 inches which was perfectly safe and usable, but not what the plaintiff had bargained for. The cost of increasing the depth of the pool was £21,460. The difference in market value between what was promised and what was received was zero.

Again we need to distinguish between the value of the contractual right infringed and loss consequent upon the infringement of that right. Costs of

[63] *Ibid*.

[64] [1995] 2 AC 207 (HL).

[65] See Law Commission, 'Privity of Contract: Contracts for the Benefit of Third Parties' (Law Comm No 242, 1996) at [2.14]; P Benson, 'Should *White v Jones* Represent Canadian Law: A Return to First Principles' in J Neyers, E Chamberlain and S Pitel (eds), *Emerging Issues in Tort Law* (Oxford, Hart Publishing, 2007) 141; Stevens, above n 5, at 176–82.

[66] [1996] AC 344 (HL) [*Ruxley*].

cure represent the expenditure which has or will be incurred as a result of the defendant's breach. This is quantified at the time of trial. It is not a quantification of the value of the right to performance at the time of breach.[67] Its only relevance to the valuation of the right at the time of breach is evidential. Where there is no ready market for the goods or services bargained for, it may provide guidance as to whether the contract price accurately reflected the right bargained for. However, the costs *ex post* of curing a breach of contract may bear no relation whatsoever to the value of the contractual right itself *ex ante*, as is demonstrated by *Ruxley*.

Claims for consequential loss, as in *Ruxley*, are subject to limitations which do not apply to claims to the value of the right. If I was summarily fired by my current employer, I would have a number of options open to me. One of those options would be to spend the rest of my life drinking beer in my local public house. If I decided to do so rather than choosing to find employment elsewhere, which employment would almost certainly be better paid than the derisory sum I currently earn, I could not recover damages representing my lost salary. Where a plaintiff unreasonably chooses to increase the loss suffered as a result of a wrong, he or she is responsible for this loss, not the defendant. The primary responsibility for the welfare of adults lies upon themselves, and this does not change where we are the victims of a wrong, whether it is a breach of contract or any other. There is no 'duty' to mitigate, but losses which ought to have been mitigated cannot be recovered.

Where therefore, as in *Ruxley*, the decision to incur the costs of cure is unreasonable, it cannot be recovered, as the House of Lords correctly held. The issue before the court was simply one of mitigation of loss.[68] We cannot, as of right, recover all losses which we would not have suffered but for the defendant's wrong. By contrast, where there is breach the difference in value between what has been promised and what has been received may be recovered as of right, regardless of reasonableness.

The House of Lords did uphold the award to the defendant of £2,500. Lord Mustill did so on the basis that 'the value of the promise to the promisee exceeds the financial enhancement of his position which full performance will secure'.[69] Lord Lloyd, by contrast, saw such damages as awarded to make good the plaintiff's consequential loss of pleasure in not having the pool he desired.[70] On the facts, the true basis for the award made no difference to the result, but the plaintiff ought to be able to recover even where he cannot show that his pleasure is any less as a result

[67] *Contra* Webb, above n 2.

[68] Burrows, above n 23, at 222.

[69] *Ruxley*, above n 66, at 360. See also *Farley v Skinner* [2002] 2 AC 732 (HL) (Lord Scott); *McAlpine v Panatown*, above n 56 (Lord Millett).

[70] *Ruxley*, *ibid*, at 373–5.

of what he received. If, for example, the plaintiff had died shortly after completion of the work his estate should still have a claim for £2,500, even though he would not have had the pleasure of using the deeper pool in any event. The value of performance is not always measured in terms of the market difference between what was promised and what was received. The value to the individual claimant, particularly in consumer contracts, may exceed this, as Lord Mustill contemplated.

VIII. BREACH OF COVENANT NOT TO BUILD

In the light of the forgoing analysis, the result in *Wrotham Park v Parkside Homes* is readily explicable, but does not have anything to do with a loss suffered by the plaintiff or a gain made by the defendant.[71] Land belonging to the defendant was subject to a restrictive covenant registered as a land charge forbidding development except with approval of the owner from time to time of the Wotham Park Estate. The defendant in breach of the covenant built 14 houses and a road on its land. The plaintiff warned the defendant of its right as soon as the work commenced, and sought an injunction restraining the work and seeking the demolition of the building which had been done. The defendant was (incorrectly) advised that the covenant was unenforceable and continued work. At trial, Brightman J refused to award the injunction, but did order damages, calculated according to a reasonable price for the claimant to have released its right.

Just as in *Semelhago*, the damages were in lieu of specific relief, and so were awarded in place of, or in substitution for, the right. Here that right was not contractual: there was no privity of contract between the land-owners. Rather, the plaintiff had a right restraining the liberty to use the neighbouring land, a right which would persist against any holder of the right to that land. As there was no ready market for quantifying this right, Brightman J calculated this value as 5 per cent of the £50,000 profit made by the defendant from the work: £2,500.

There was no loss suffered. The value of the plaintiff's land was not reduced by the defendant's building. In some cases it might be argued that the relevant loss is the lost opportunity to bargain for the release of the covenant,[72] but as Andrew Burrows pointed out 20 years ago, this will not be realistic where the defendant would not have agreed to any such release, as was the case in *Wrotham Park*.[73]

[71] [1974] 1 WLR 798 (Ch D).

[72] R Sharpe and S Waddams, 'Damages for Lost Opportunity to Bargain' (1982) 2 *OJLS* 290.

[73] Burrows, above n 23, at 275.

It does not assist to describe the claim as being made in order to compensate for an 'objective' or 'normative' loss. All that such labels show is that the 'loss' is not actual or real but rather fictional or deemed. Unless the claimant is, as things have turned out, in some way factually worse off (although not necessarily economically) he has suffered no loss. Worse, such expansive use of 'loss' is dangerously loose as it misleads us into thinking that damages, which are substitutive for the right, are compensatory and are therefore subject to the same principles as apply to damages which are awarded to make good consequential losses. As we have seen, however, the rules on remoteness, mitigation, incidental benefits, timing of assessment and quantification are all different for the two heads of damage. We should give different labels to different things, especially where consequences turn upon the difference.

Even worse is the attempt to classify the claim as gain-based. First, the claim is not quantified according to the actual gain made by the defendant, as is illustrated by *Wrotham Park* itself. The actual profits or expenses saved do not form the basis of the award, save in so far as they evidence the value of the right. Second, in a truly gain-based award the defendant's overall position as a result of the wrong at the time of trial would be examined. Losses which are suffered as a result of the wrong would be capable of being offset against the gains made. In fact, no such offsetting is possible, so that a claim for substantial damages would succeed even where defendants make overall losses from the wrong.[74] Third, describing the claim as one to the 'normative' or 'objective' gain is as fictitious as describing the claim as one to the 'normative' or 'objective' loss. As a matter of language, it is possible to describe the infringement of a right as, in itself, a 'loss', although as a matter of law this conflation of *injuria* and *damnum* is always unhelpful. To describe the infringement as per se a gain to the defendant is inaccurate, even as a matter of language. The right is not, as has been claimed,[75] transferred to the defendant. The right itself has not been 'gained', it is still vested in the plaintiff.[76] Fourth, the popularity of trying to classify all claims for damages which are not punitive (and hence considered anomalous) as either based upon the plaintiff's loss or the defendant's gain has resulted in the dismissal as incorrect all cases which cannot plausibly be so classified. So, for example, Burrows considers *The Mediana*, *William Bros*, *Rodocanachi Sons*, *Slater*, *Campbell Mostyn*,

[74] *Inverugie Investments Ltd v Hackett* [1995] 1 WLR 713 (PC); See Stevens, above n 5, at 79–84.

[75] J Edelman, *Gain-Based Damages: Contract, Tort, Equity and Intellectual Property* (Oxford, Hart Publishing, 2002) 117–18.

[76] C Rotherham, 'The Conceptual Structure of Restitution for Wrongs' [2007] *CLJ* 172; Stevens, above n 5, at 80–81.

Jamal, *Joyner v Weeks,* and *Semelhago v Paramadevan* all to be wrongly decided, and the minority views of Lords Jauncey and Clyde in *McAlpine v Panatown* to be preferable.

Fifthly, and most dangerously, if cases such as *Wrotham Park* are misclassified as gain-based it is a short step from there to saying that when the defendant has made an *actual* gain from the wrong committed, he should be required to disgorge it to the plaintiff. However, once the rights-based analysis defended here is accepted, the category of gain-based damages is rendered otiose save where the actual gain made by the defendant is greater both than the value of the right infringed and the consequential loss suffered. It is difficult to identify any cases, prior to the decision of the House of Lords in *Attorney General v Blake*,[77] which permitted such recovery for breach of a contract, or indeed a tort.[78] Analogies with the liability of fiduciaries to account for profits does not assist. It is, at the lowest, arguable that such liability is better classified as a primary rather than a secondary response to wrongdoing. It is one of the primary duties of fiduciaries that they are obliged to account for the profits they make in that capacity. Part of the duty to subordinate your interests to those of someone else is that you must account for profits you make out of the relationship.[79] Alternatively, in some cases the duty to account for profits may be seen as substitutive for the right of the beneficiary which has been infringed.

Lord Nicholls, in delivering the majority speech in *Blake*, described *Wrotham Park* as a 'solitary beacon'[80] permitting the award of gain-based damages for breach of contract. As the claim in *Wrotham Park* was not based upon the defendant's gain, nor indeed was it for breach of contract, it is difficult to resist the conclusion that the beacon caused their Lordships to hit the rocks, as Lord Hobhouse forcefully pointed out in his dissent.

IX. NOMINAL DAMAGES

It has been objected that the account of the law given above leaves no room for the award of nominal damages.[81] This is not so.

First, the right infringed may be valueless. If, for example, I was summarily fired from my job without justification, only nominal damages

[77] *A-G v Blake*, above n 47.

[78] On the absence of genuinely gain based awards for torts, see Stevens, above n 5, at 83–4.

[79] P Millett, 'Equity's Place in the Law of Commerce' (1998) 114 *LQR* 214, 225–7.

[80] *A-G v Blake*, above n 47, at 283.

[81] See A Burrows, 'Are "Damages on the *Wrotham Park* Basis" Compensatory, Restitutionary, or Neither?' in R Cunnington and D Saidov (eds), *Contract Damages: Domestic and International Perspectives* (Oxford, Hart Publishing, 2008).

would be available to me at common law.[82] Although I have suffered a wrong, my right to contractual counter-performance from my employer is essentially valueless. My right to be paid my salary is conditional. In order to earn it, I must work. The work I provide, in terms of teaching, research, supervision and administration is objectively worth far more than the pittance I am paid, as all but the most obtuse observer would accept. Unless I could prove consequential loss, no substantial damages would be payable as my right to be paid at the end of each month is worth less than the work I must do each month in order to get it.

Second, a valuable right may be infringed in an insignificant way. If you scratch my Rolls Royce you obviously need not pay me the full value of the car. What is quantified is the infringement of the right. If the right is valueless or the infringement is notional only nominal damages are payable.

X. *THE GOLDEN VICTORY*

The decision of the House of Lords in *Golden Strait Corporation v Nippon Yusen Kubishika Kaisha (The Golden Victory)*[83] may at first sight appear to be inconsistent with the law as I have so far described it.[84] Indeed, the speeches for the majority emphasise that damages for breach of contract are compensatory for actual loss suffered.[85] In fact this decision nicely illustrates the difference between damages in relation to performance which has accrued, and those in respect of future performance.

The defendants chartered the plaintiff owner's vessel under a time charter. In repudiation of the contract, they redelivered the vessel when the contract still had four years left to run. The owners accepted the repudiation and claimed damages, calculated by reference to the entire four years that the contract still had outstanding. The defendants sought to reduce the damage payable on the basis that 14 months after the repudiation was accepted, the second Gulf War broke out, which would have given both parties the option to lawfully terminate the contract, an option that the charterers would have exercised at that point. The defendants, therefore, argued that as no loss was suffered by the owners after the outbreak of war, no damages should be recoverable after it commenced. The plaintiffs

[82] I would, in addition to a claim for breach of contract for wrongful dismissal, also have a statutory claim for unfair dismissal: Employment Rights Act 1996.

[83] [2007] 2 AC 353 (HL) [*The Golden Victory*].

[84] Indeed, on the Obligations Discussion Group I rashly said I thought it was wrong: see <http://www.ucc.ie/law/odg/messages/070328b.htm> and the subsequent discussion (last accessed: 8 August 2008). Mea culpa.

[85] *The Golden Victory*, above n 83 (Lord Scott, Lord Carswell and Lord Brown; Lord Bingham Cornhill and Lord Walker, dissenting).

sought to argue that damages should be assessed at the time of breach. Their argument was that the risk of a future war affected the market value of the charter at that point, but as war was then far less certain, substantial damages should be awarded representing the entire value of the charter at time of breach. By a bare majority, the House of Lords found for the charterers, upholding the decision of the Court of Appeal.

The minority would have taken the time of breach as the time at which damages were crystallised based upon considerations of certainty and finality. The quantum of damages would otherwise vary according to the chance of the moment at which judgment was given.[86] Such an argument, if truly accepted, would lead to the conclusion that events occurring after the time of breach should never be taken into account in assessing damages. This is clearly not the law. Certainty is a second order principle of justice. A rule which said that the damages for breach of contract are always fixed at $100 would be even more certain, but nobody would accept it as an acceptable rule.

The majority emphasised that damages are compensatory for actual loss suffered, and to ignore intervening events which have reduced this loss would lead to over-compensation. If, however, this is accepted, how are the cases where damages are assessed at the time of breach, ignoring subsequent events, to be explained?

The answer is that *The Golden Victory* was a claim with respect to obligations to perform which had not accrued due at the time of repudiation. Exceptionally, the ability to claim damages can be accelerated. Once a repudiation has been accepted, damages for the loss of future performance can be claimed at the time of acceptance, even though the time for such performance has not yet fallen due. However, what is accelerated is the ability to claim damages, not the right to future performance itself. The value of the right to performance must be quantified at the time performance is due. In cases of the breach of an accrued contractual right, the right is valued at the time of breach. Consequential losses are assessed at time of judgment. Where damages are awarded *in lieu* of a specific remedy, the right to performance is quantified at the time of trial. Where damages are claimed with respect to future performance, the contractual right must be valued at the time at which it would accrue to be performed. In *The Golden Victory,* it was known that after the outbreak of the second Gulf War, the owner's rights to further performance from the charterers was valueless because of the right to terminate. Therefore, regardless of whether the claim is seen as one for consequential loss or to the value of the charter, damages should have been and were discounted to take

[86] See also GH Treitel, 'Assessment of Damages for Wrongful Repudiation' (2007) 123 *LQR* 9.

account of the outbreak of war. The owners would never have had a valuable right to performance of the charter after the outbreak of war.

XI. CONCLUSION

In law, the purchase of my wife's wedding ring has a happy conclusion. Substantial damages are payable. This is not because of my 'reliance' upon the jeweller's promise. This may be illustrated by some (fictional) figures. The contract price for the ring as promised was £500, but the market price was £2,000 (that is, I made a good deal, or at least I thought I did). The plated ring delivered was only worth £50. My 'reliance' was no greater than the £500 I paid. The damages payable are £1,950, which is not equivalent to any sum I overpaid.

Similarly, the claim is not equivalent to a claim measured by the amount that the seller made from my payment. This gain is not £1,950. The gain made is £450, the difference between what was paid and what the ring is actually worth. A gain-based analysis does not explain the law either.

Whether the right to performance *should* be treated sufficiently seriously that its violation per se gives rise to a right to damages is a more difficult question to answer. Some rights are treated by the law as more important, and consequently more deserving of respect, than others. Some wrongs are actionable per se, some require the proof of consequential loss and some are not actionable at all. The breaking of promises supported by consideration falls into the first category, lies fall into the second, and insults fall into the third. Some rights are more important than others, and consequently some wrongs are worse than others. However, there is no mathematical formula or proof that enables us to demonstrate such ranking in the abstract. In the end, we are each of us on our own in judging the importance which ought to (morally) be attached to different rights. It *could* be the law that we treated the right to contractual performance in the same way as our right not to be slandered, so that only consequential losses gave rise to a claim for damages. Indeed it *could* be the law that contractual rights and their infringement were not actionable in law at all, just like insults.[87] The law is usually irrelevant in most contractual disputes, as anyone who has got into a dispute with a builder will readily understand. Gaming contracts were for many years void in England,[88] but this did not inhibit the enormous gaming industry. Betting companies very rarely fail to pay out, unlike insurance companies who are in essentially the same business but whose contracts are recognised by the law. Commerce is less dependent upon the law of contract than lawyers like to believe.

[87] In Scotland, insults are actionable, as they were in Roman law.
[88] Gaming Act 1845, s 18; but see now Gaming Act 2005 ss 334–5.

At the margins, the law is there to provide determinate rules where morality is underdetermined. It may not be demonstrable whether the choice the common law has made is correct or incorrect. Whatever the merits of different choices, there is too much law to the contrary for it to be seriously argued that the law of damages for breach of contract is solely concerned with compensating losses (or stripping of gains). This has been hidden from us for two reasons. First, in the majority of cases, but not all, the loss suffered is either greater than or the same as the value of the right. In travelling to the symposium at which this article was delivered my baggage was lost by Air Canada. I had tempted fate by writing about the non-delivery of goods. Damages calculated by reference to the market value of my second-hand suits and shirts would clearly be inadequate. It is reasonable for me to incur the higher costs of replacement, and this will be the usual sum awarded. Care must be taken that this normal position does not obscure the entitlement to damages even where such consequential loss is not incurred.

Second, misleading terminology has been adopted, with terms such as 'compensation', 'loss' and 'restitution' being stretched to cover quite different ideas with different incidental rules.

The first task of the academic lawyer is to explain the law so that it makes coherent sense and to account for it in the best possible light. In the common law, we treat the right to contractual performance as sufficiently important that when it is infringed, it gives rise to a claim for damages representing its value. My judgement, for what it is worth, is that this is (morally) appropriate. My wife agrees.

8

Estoppels and Rights-Creating Events: Beyond Wrongs and Promises

ANDREW ROBERTSON*

THIS ARTICLE AIMS to explain the nature of the substantive doctrines of estoppel as a source of rights and obligations. The main focus is on proprietary estoppel because that doctrine operates as a substantive, independent source of rights in several jurisdictions including England, Australia and Canada. The analysis extends to the broader Australian doctrine of equitable estoppel, which is essentially the same set of principles freed from the limitation that the assumption relied upon must relate to an existing or expected interest in property. The expression 'equitable estoppel' will be used to encompass both proprietary estoppel and the Australian doctrine, which is broader in scope than proprietary estoppel but precisely the same in principle. The doctrines will be analysed by considering their place within a taxonomy of the law of obligations organised around rights-creating events. This is an exercise in interpretative legal theory,[1] rather than moral philosophy, and so its focus is on explaining the pattern of liability in the case law rather than identifying a moral foundation for that liability.[2]

The doctrine of proprietary estoppel merits more attention than it receives in theoretical literature on contract law.[3] Proprietary estoppel is an

* I would like to thank Ben McFarlane and the participants in the 'Exploring Contract Law' symposium for their helpful comments on earlier drafts of this article.

[1] S Smith, *Contract Theory* (Oxford, Oxford University Press, 2004) 5–6; A Beever and C Rickett, 'Interpretive Legal Theory and the Academic Lawyer' (2005) 68 *MLR* 320, 321–5.

[2] Whether the doctrines make 'moral sense' or whether a different pattern of liability might make better moral sense are therefore outside the scope of this paper: see Smith, *ibid*, at 243–4. A moral justification of the obligation created by a substantive doctrine of estoppel has been advanced by M Spence, *Protecting Reliance: The Emergent Doctrine of Equitable Estoppel* (Oxford, Hart Publishing, 1999) ch 1.

[3] See, eg, Smith, above n 1, at 233–4.

important source of rights arising from dealings that often involve promises, and even exchanges of promises. The doctrine is regularly applied by English and Australian courts in a wide range of contexts.[4] It has been accepted that similar principles operate in Canada, although the doctrine seems to be applied far less frequently by Canadian courts.[5] The English doctrine of promissory estoppel, in contrast, receives an amount of attention in the literature on contract law and contract theory that is perhaps undeserved, since the doctrine has a limited scope of application and is applied by the courts only occasionally.[6] As Susan Bright and Ben McFarlane have observed, the notion that estoppel provides a cause of action only in property cases has no obvious justification and has persistently been criticised.[7] A rationalisation is overdue.[8] Until that occurs, however, the scholar interested in reliance-based liability arising from non-contractual exchange transactions, gratuitous promises and other contexts in English law must turn to proprietary estoppel.

In England and Australia, proprietary estoppel is regularly applied both inside and outside the familiar contexts of dealings between family members in relation to residential and commercial property, promises of property made to caregivers and dealings between neighbours about rights of way. The doctrine has also recently been applied, for example, in cases involving dealings between creditors concerning the priority of charges and

[4] A recent study identified and discussed 18 English and 19 Australian cases over a five-year period in which remedies were granted to give effect to proprietary estoppel: A Robertson, 'The Reliance Basis of Proprietary Estoppel Remedies' [2008] *The Conveyancer and Property Lawyer* 295.

[5] See, eg, *Romfo v 1216393 Ontario Inc* (2007) 285 DLR (4th) 512 (BCSC).

[6] A search of the 'UK Cases Combined' LexisNexis database (on 28 November 2007) for cases in the preceding five years in which the words 'promissory estoppel' were used identified only three cases in which the principle of promissory estoppel was applied: *Msas Global Logistics Ltd v Power Packaging Inc* [2003] EWCH 1393; *Bottiglieri Di Navigazione SpA v Cosco Qingdao Ocean Shipping Company (The 'Bunga Saga Lima')* [2005] 2 Lloyd's Rep 1 (QB); *Taylor v Rive Droite Music Ltd* [2004] EWHC 1605 (Ch). In the last two of those cases promissory estoppel was an alternative justification for results reached on the basis of, respectively, collateral contract and interpretation of the contract terms. The principle was also applied in *Business Environment Bow Lane Ltd v Deanwater Estates Ltd* [2006] EWHC 3363 (Ch), but that decision was overturned by the Court of Appeal: [2007] EWCA Civ 622.

[7] S Bright and B McFarlane, 'Personal Liability in Proprietary Estoppel' [2005] *The Conveyancer and Property Lawyer* 14. As to the current position in English law, see *Baird Textile Holdings Ltd v Marks & Spencer plc* [2002] 1 All ER (Comm) 737 (CA).

[8] See J Neyers, 'A Coherent Law of Estoppel?' (2003) 2 *Journal of Obligations and Remedies* 25, 31–3; D Nolan, 'Following in their Footsteps: Equitable Estoppel in Australia and the United States' (2000) 11 *King's College Law Journal* 202; B McFarlane and D Nolan, 'Remedying Reliance: The Future Development of Promissory and Proprietary Estoppel in English Law' (paper presented at the Obligations III conference, University of Queensland, July 2006) (cited with authors' permission); B McFarlane, 'The Problem of Pre-Contractual Reliance: Three Ways to a Third Way' (2006) *Hauser Global Law Working Paper* 12/06 (http://www.law.nyu.edu/global/workingpapers/2006/ECM_DLV_015762) (last accessed November 2008).

dealings between liquidators and company directors concerning the ownership of property.[9] The doctrine is also being applied to property other than land. In a recent English case proprietary estoppel was held to arise from dealings between partners in an accounting firm concerning beneficial interests under a trust of a life insurance policy.[10] In the well-publicised litigation relating to the copyright in the song A Whiter Shade of Pale it was assumed that the claimant could be prevented from asserting his copyright interest in the song by way of proprietary estoppel (although ultimately he was not).[11] Moreover, Lord Hoffmann recently observed that 'There is no reason why the equitable rules of proprietary estoppel should not apply to a patent in the same way as to any other property.'[12] The broader Australian doctrine has been applied in a range of non-proprietary contexts, although interestingly the great majority of the recent Australian cases involving the application of equitable estoppel have fallen within the scope of the orthodox doctrine of proprietary estoppel.[13]

I. WRONGS AND RIGHTS-CREATING EVENTS

Peter Birks has argued that the law of obligations can usefully be understood by reference to four categories of rights-creating events: '(1) contracts (consent), (2) torts (wrongs), (3) unjust enrichments, and (4) other events'.[14] There are two significant and related problems with the category of wrongs. The first is that the category is too broad and abstract

[9] *Scottish & Newcastle plc v Lancashire Mortgage Corporation Ltd* [2007] EWCA Civ 684; *Roufeil v Lusby* [2003] NSWSC 1002; *Hypec Electronics Pty Ltd v Mead* (2003) 202 ALR 688 (NSWSC), aff'd [2004] NSWCA 221.

[10] *Strover v Strover* [2005] EWHC 860 (Ch).

[11] *Fisher v Brooker* [2008] EWCA 287.

[12] *Yeda Research and Development Co Ltd v Rhône-Poulenc Rorer International Holdings Inc* [2008] 1 All ER 425 (HL) [22].

[13] A search of the Lexis-Nexis 'Australian Commonwealth, State and Territory Caselaw' database for cases over a five-year period from July 2002 to July 2007 identified only five cases falling outside the scope of proprietary estoppel in which remedies were granted to give effect to equitable estoppel: *EK Nominees Pty Ltd v Woolworths Ltd* [2006] NSWSC 1172 (involving dealings between a landlord and a prospective tenant of a supermarket); *Anaconda Nickel Ltd v Edensor Nominees Pty Ltd* [2004] VSCA 167 (involving dealings between a proposed buyer and seller of shares); *New Zealand Pelt Export Company Ltd v Trade Indemnity New Zealand Ltd* [2004] VSCA 163 (involving dealings between a trader and trade indemnity insurer concerning the application of the policy to sales not covered by its terms); *GEC Marconi Systems Pty Ltd v BHP Information Technology Pty Ltd* (2003) 128 FCR 1 (involving dealings between parties to a commercial contract involving a contract variation); *Gray v National Crime Authority* [2003] NSWSC 111 (involving dealings between the National Crime Authority and witnesses entering the witness protection program relating to financial arrangements).

[14] P Birks, 'The Concept of a Civil Wrong' in DG Owen (ed), *Philosophical Foundations of Tort Law* (Oxford, Clarendon Press, 1995) 31, 33.

to have any explanatory power.[15] For that reason, an obligation should be regarded as falling within the category of wrongs only if it cannot properly be understood as falling within one of the other categories. The second problem is that while consent and unjust enrichment are events that give rise to primary rights, the category of wrongs is concerned, anomalously, with events that give rise to secondary rights. Birks accepted this anomaly because, for the obligations that fall within this category, it is not possible to identify at a generic level the events that give rise to the primary rights.[16] Robert Stevens has offered a classification of the law of torts at the level of primary rights—the right to bodily safety, freedom of movement, reputation, and so on—but suggests that the fact that many of these rights are generated by the event of birth demonstrates the limited utility of an events-based categorisation.[17]

Although the Birksian events-based scheme has significant limitations, particularly when it comes to understanding the law of torts, it does provide a useful framework for analysing a substantive doctrine of estoppel (that is, one that operates as an independent source of rights). The questions raised by the events-based taxonomy go to the heart of equitable estoppel, and a consideration of those questions provides significant insights. Equitable estoppel can potentially be placed in three categories within the events-based taxonomy. First, it could be classified within the law of wrongs. Whether equitable estoppel is properly classified within the law of wrongs depends on a negative answer to the question of whether it is possible to identify an event, or series of events, which gives rise to primary rights that are recognised by the law prior to and independently of any infringement of those rights.[18] If it is possible to identify an event that gives rise to primary rights, then estoppel does not belong within the law of wrongs. It should be classified by reference to the event that gives rise to the primary rights. Second, equitable estoppel could be classified, as Birks suggested, as a consent-based obligation.[19] This depends on whether the basis of obligation under a substantive doctrine of estoppel is the making of a promise or some other manifestation of consent. Third, equitable estoppel could be classified as an obligation that arises from an 'other' rights-creating event. If equitable estoppel gives effect to primary rights that are created by an event other than consent, then equitable estoppel must be regarded as a *sui generis* source of rights falling, in the Birksian scheme, within the 'other' category of rights-creating events.

[15] *Ibid*, at 51.
[16] *Ibid*, at 46.
[17] R Stevens, *Torts and Rights* (Oxford, Oxford University Press, 2007) 287. *Cf* Birks, above n 14, at 47.
[18] Birks, above n 14, at 50–51.
[19] P Birks, 'Equity in the Modern Law: An Exercise in Taxonomy' (1996) 26 *University of Western Australia Law Review* 1.

A. Inconsistent Conduct as a Wrong

I have argued elsewhere, incorrectly, that equitable estoppel is best understood as part of the law of wrongs.[20] That argument was based on the idea that a wrong is committed when A, having induced B to adopt and act upon an assumption, acts inconsistently with the assumption without taking steps to ensure that B is not harmed as a result of that action.[21] Many equitable estoppel cases are brought on the basis of a threatened departure from the assumption in question, rather than one that has actually been committed. These cases can be reconciled with the view that equitable estoppel deals with wrongs, on the basis that equitable estoppel operates in some cases to restrain apprehended wrongs and in others to redress wrongs already committed. On this view, equitable estoppel has a significant *quia timet* jurisdiction.

The crucial issue for an events-based taxonomy, however, is not whether equitable estoppel responds to conduct that can be characterised as a wrong. Rather, we must ask whether it is possible to characterise the relevant rights as arising from something other than the commission of a wrong. The fundamental question is whether it is possible to identify an event, or series of events, that creates primary rights recognised by the law prior to and independently of any infringement of those rights. The notion that equitable estoppel is concerned with 'unconscionable conduct' has led to a view that no identifiable rights arise by way of estoppel until there has been unconscionable conduct. In *Ashton Mining Ltd v Commissioner of Taxation*, Merkel J held that 'the rights created by an equitable estoppel cannot arise until there has been an unjust or unconscionable departure or threat to depart from the assumption adopted and acted upon by the party seeking to assert the estoppel'.[22] This conclusion was based on two closely related ideas that are commonly said to be foundational to estoppel. The first is that the purpose of estoppel is to prevent an unjust departure from an assumption which has been acted upon. The second is that 'the element which both attracts the jurisdiction of a court of equity and shapes the remedy to be given is unconscionable conduct on the part of the person bound by the equity'.[23] The purpose of estoppel and the equitable jurisdiction are not, on this view, enlivened unless there is a threat of unconscionable conduct in the form of an unjust departure from the

[20] A Robertson, 'Situating Equitable Estoppel within the Law of Obligations' (1997) 19 *Sydney Law Review* 32, 57–60.

[21] A Robertson, 'Reliance, Conscience and the New Equitable Estoppel' (2000) 24 *Melbourne University Law Review* 218, 227.

[22] [2000] ATC 4307 (FCA) [50] [*Ashton Mining*].

[23] *Waltons Stores (Interstate) Ltd v Maher* (1988) 164 CLR 387 (HCA) 419 (Brennan J) quoted in *Commonwealth v Verwayen* (1990) 170 CLR 394 (HCA) 411–12 (Mason CJ) and *Sledmore v Dalby* [1996] 72 P & CR 196 (CA) 208.

relevant assumption. Thus, in *Commonwealth v Verwayen*, Dawson J said that: 'An estoppel will occur only where unconscionable conduct on the part of one gives rise to an equity on the part of another.'[24] This is consistent with Kevin Gray and Susan Francis Gray's view that the 'inchoate equity' arising by way of proprietary estoppel 'is brought into being when the landowner unconscionably sets up his rights adversely to the legitimate demands of the estoppel claimant'.[25]

The issue in *Ashton Mining* was whether Ashton Mining Ltd (AM) was entitled to a tax deduction with respect to a debt that AM had 'written off'. The debt had been owed to AM by its subsidiary, Ashton Gold Ltd (AG). A deduction was allowable under the Income Tax Assessment Act if 'all liability to pay any amount or amounts' had been discharged.[26] Whether liability to pay the debt had been discharged in this case depended, inter alia, on whether AM 'by its conduct, was estopped from claiming that the Ashton Gold debt had not been discharged'.[27] Merkel J did not accept that AG had adopted or acted upon an assumption that the debt had been discharged.[28] In the end, however, he rested his decision on the 'additional difficulty with the estoppel case',[29] which was that 'There was no departure, or threatened departure, by Ashton Mining from any assumption adopted by Ashton Gold during the 1992 year of income.'[30] The decision was therefore based on the notion that, without threatened or actual inconsistent conduct causing harm, no rights arise by way of estoppel.

The question of principle that arises from this decision is whether, if A has induced B to adopt an assumption or expectation as to A's future conduct or B's legal rights, which B has acted upon (in such a way that B would suffer detriment if A behaved inconsistently with the assumption), are the rights and obligations of the parties affected by that course of events if A has not departed or threatened to depart from the assumption? The issue becomes particularly clear if one takes the familiar family property example, where A (the property owner) promises B (her son) that A will allow B and his family to reside in a particular house owned by A and that A will leave this property to B in her will. B and his family invest

[24] *Ibid*, at 453–4.

[25] K Gray and SF Gray, *Elements of Land Law*, 4th edn (Oxford, Oxford University Press, 2005) 967–8. *Cf* B McFarlane, 'Proprietary Estoppel and Third Parties after the Land Registration Act 2002' [2003] *CLJ* 661.

[26] Income Tax Assessment Act 1936 (Cth) ss 70B(2) and 159GP(1).

[27] Above n 22, at [43].

[28] It is interesting to note, however, that the following year the shares in AG were purchased by a third party on the assumption that the debt had been discharged, so in slightly different circumstances the estoppel issue could have been a live one, even without any attempt by AM to depart from the assumption.

[29] Above n 22, at [45].

[30] *Ibid*, at [54].

considerable personal effort and funds in renovating the house, reside in it for 10 years upon the assumption that the house is theirs, and develop a significant emotional attachment it. A has no intention of acting inconsistently with the assumption B has adopted. For the purpose of determining A's pension entitlements, however, should A be regarded as the sole beneficial owner of the property?

Once B has taken (potentially detrimental) action upon an assumption induced by A, there can be no doubt that A's freedom of action is constrained. The estoppel already governs the relationship between the parties because A is not, at this point, free to act inconsistently with the assumption that she has induced B to adopt and act on, at least without taking steps to ensure that departure from the assumption does not cause harm to B. As Ben McFarlane has argued, B must therefore be regarded as having rights against A by way of estoppel from the time reliance occurs.[31] It follows that departure or threatened departure from the assumption cannot be regarded as a prerequisite to an interest in land arising by way of proprietary estoppel. On the facts of *Ashton Mining*, assuming some reliance by the debtor, it makes no sense to say that the creditor's right to recover the debt remains unimpaired until the creditor attempts to exercise that right, but the creditor will be prevented from recovering the debt if it attempts to do so. If the creditor will be prevented from recovering the debt if it attempts to do so, then its right to the debt is clearly impaired before it so attempts.

This idea has now been recognised in several Australian cases.[32] It has been held in these cases that, for the purpose of determining pension entitlements, a beneficial interest in land can effectively pass by way of equitable estoppel in the absence of any attempt by the landowner to resile from the relevant assumption. In *Repatriation Commission v Tsourounakis*,[33] for example, Mr and Mrs Tsourounakis owned a house property which they allowed their son and his family to occupy. The son was induced to believe that he could occupy the property rent-free and consider it his own and that the parents would leave the property to him in their wills. The son and his family spent 'a large amount of time, energy and money repairing and renovating the Property' on the assumption that it was theirs.[34] Many years later, the question arose whether Mr and Mrs Tsourounakis should be regarded as the beneficial owners of the property

[31] B McFarlane, 'Blue Haven Enterprises Ltd v Tully & Another' (2006) 1 *Journal of Equity* 156, 159, building on McFarlane, above n 25, at 665–74.

[32] In addition to the cases discussed in the text, see *Kintominas v Secretary, Department of Social Security* (1991) 103 ALR 82 (FCA); *Sarkis v Deputy Commissioner of Taxation* [2005] ATC 4205 (VCA); *Secretary, Department of Family and Community Services v Wall* [2006] FCA 863.

[33] [2004] FCAFC 332.

[34] *Ibid*, at [50].

for the purpose of determining their entitlement to a veteran's service pension. The Repatriation Commission argued that since Mr and Mrs Tsourounakis had made no attempt to resile from the assumption, the son had 'no cause of action that would result in his being recognised as having any equitable claim or interest in respect to the Property.[35] Spender, Kiefel and Emmett JJ held:

> That contention is somewhat facile. The question is not whether Mr and Mrs Tsourounakis are threatening to act in an unconscionable manner. The question is whether, if they did, Michael would be entitled in equity to restrain them from doing so. If he would, the value of the Property to Mr and Mrs Tsourounakis must be diminished to the extent that they would be required to compensate Michael as a term of avoiding any restraint by a court of equity. If a court of equity would treat Michael as the beneficial owner of the Property, the value of the interest of Mr and Mrs Tsourounakis must be regarded as nil.[36]

The question whether equitable estoppel creates rights without inconsistent conduct is not only relevant for determining issues relating to taxation and social security entitlements. The question can have implications for the rights and obligations asserted between individuals where the inchoate estoppel is relied upon by a third party. This situation arose in *Young v Lalic*.[37] Mrs Lalic induced her son (Mr Lalic) to believe that if he built a house on land she owned, she would allow him to reside in the house indefinitely and would ultimately convey it to him. The son induced his fiancée, Ms Young, to believe that the mother would permit Mr Lalic and Ms Young to occupy the property together and would ultimately transfer it to them. On the faith of this assumption Ms Young paid $50,000 into the bank account of Mrs Lalic for the purpose of financing the construction of the house. The relationship between the son and Ms Young foundered before the house was complete. Although Mrs Lalic did not ever act inconsistently with the assumption she induced her son to adopt, Ms Young sought relief against her. Brereton J held that Ms Young could not assert an estoppel against Mrs Lalic because Mrs Lalic did not induce Ms Young to adopt the relevant assumption and did not know the money had been paid into her bank account. But Brereton J held that Mr Lalic had acquired an interest in his mother's land by way of equitable estoppel, and that interest could be conveyed, or pass by way of equitable estoppel, to a third party.[38] Mr Lalic had, by his conduct, conferred 'a subsidiary equitable interest on Ms Young, which she can enforce directly against Mrs Lalic'.[39]

[35] *Ibid*, at [51].
[36] *Ibid*.
[37] (2006) 197 FLR 27 (NSWSC).
[38] *Ibid*, at [83], citing *Hamilton v Geraghty* (1901) 18 WN (NSW) 152 (SC).
[39] *Young v Lalic*, above n 37, at [83].

Thus, although Mrs Lalic had not behaved inconsistently before Ms Young instituted proceedings,[40] she had induced her son to adopt an assumption that the land would be conveyed to him. His reliance upon that assumption gave him an interest in the land by way of estoppel. He in turn was able to confer 'a subsidiary equitable interest' in the property on Ms Young by way of estoppel. Ms Young was able to rely on Mr Lalic's estoppel against Mrs Lalic as a basis for establishing her own claim against Mrs Lalic. Although Brereton J did not discuss the fact that Mrs Lalic had not behaved inconsistently before the time the proceedings were instituted, the recognition that Mr Lalic had rights against his mother by way of estoppel, and an interest in her land, is consistent with the decisions discussed above, and is justified for the same reasons.

The Australian pension cases discussed above, along with *Young v Lalic*, show that a series of events consisting of assumption, inducement and action in reliance gives rise to primary rights which are recognised by the law before the inducing party has engaged in any inconsistent conduct. These decisions must be right as a matter of principle. Once there has been detrimental reliance by one party upon an assumption induced by another it is clear that the inducing party's freedom of action is constrained. The inducing party has an obligation to the relying party and the relying party has corresponding rights against the inducing party. The estoppel governs the relationship between the parties before the inducing party engages in any inconsistent conduct. If inconsistent conduct causing harm is regarded as the wrongful conduct in proprietary estoppel, then it is clear that proprietary estoppel is not part of the law of wrongs. That is because identifiable, legally-recognised primary rights pre-date the commission of any wrong.

B. Inducing a Change of Position as a Wrong

The analysis above depends on the notion that it is inconsistent conduct that constitutes the wrong. Proprietary estoppel could, however, be regarded as part of the law of wrongs if the inducement of a change of position, rather than inconsistent conduct, was regarded as the wrong. Peter Benson has suggested that 'a reliance-based action for breach of promise should be recognized and approached as one variant of tort

[40] In the proceedings brought by Ms Young, Mrs Lalic denied telling Mr Lalic that she would transfer the land to him and 'said that she did not propose to transfer land to Mr Lalic': *ibid*, at [93].

liability for negligent statements or representations'.[41] This follows Warren Seavey's view that the wrong in promissory estoppel is not in disappointing the claimant's expectation, but in causing his or her change of position:

> Estoppel is basically a tort doctrine and the rationale of [section 90] is that justice requires the defendant to pay for the harm caused by foreseeable reliance upon the performance of his promise. The wrong is not primarily in depriving the plaintiff of the promised reward but in causing the plaintiff to change position to his detriment.[42]

This understanding, if correct, would square the idea that estoppel is based on a wrong with the notion that rights arise by way of estoppel before any inconsistent conduct. The crucial question, however, is whether causing the claimant to change his or her position, in circumstances in which that reliance is foreseeable, can be regarded as wrongful. The reason it cannot is that there is nothing wrong, in itself, with inducing an assumption upon which you know another person will rely. Provided the assumption is made good, the relying party suffers no harm as result of his or her reliance. The relying party may even benefit in some situations.[43] As Charles Goetz and Robert Scott have pointed out, a gratuitous promise gives the promisee a valuable piece of information about the future, which the promisee can use to his or her advantage by acting on the promise.[44] If the promisee's expectations are fulfilled, the promisee stands to benefit from the advance knowledge. Goetz and Scott called this 'beneficial reliance'. This insight makes it clear that inducing reliance cannot in itself be regarded as wrongful. It can only be considered wrongful if there is some reason to doubt the reliability of the assumption, and thus some prospect of harm to the relying party.

If negligent conduct is to be understood as the basis of equitable estoppel, therefore, it would need to consist of a failure to warn the claimant of the risk that the assumption might not be fulfilled. The defendant's failure to warn the claimant can be seen as careless if a reasonable person would have foreseen the circumstances which caused the defendant to change his or her mind and would have taken steps to warn the claimant about that risk. It is possible that some equitable estoppel cases can be understood this way. In *Waltons Stores (Interstate) Ltd v*

[41] P Benson, 'The Expectation and Reliance Interests in Contract Theory: A Reply to Fuller and Perdue' [2001] *Issues in Legal Scholarship* 5, 61. See also P Benson, 'The Unity of Contract Law' in P Benson (ed), *The Theory of Contract Law: New Essays* (Cambridge, Cambridge University Press, 2001) 118, 177.

[42] WA Seavey, 'Reliance upon Gratuitous Promises or Other Conduct' (1951) 64 *Harvard Law Review* 913, 926.

[43] For an example, see A Robertson, 'The Failure of Economic Analysis of Promissory Estoppel' (1999) 15 *Journal of Contract Law* 69.

[44] C Goetz and R Scott, 'Enforcing Promises: An Examination of the Basis of Contract' (1980) 89 *Yale Law Journal* 1261.

Maher,[45] for example, one might be able to say that a reasonable person in the position of the prospective tenant would have foreseen the risk that, because the tenant was reconsidering its retailing strategy, the lease transaction may not go ahead, and would have taken steps to warn the prospective landlord of this risk before the landlord went ahead with building work.

This, however, opens up factual questions that the courts do not consider in the estoppel cases. Whether the defendant's conduct can be regarded as careless must depend on what the defendant knew about the circumstances that might (and ultimately did) lead the defendant to disappoint the claimant's expectations, and what inquiries a reasonable person would have made. In quite a number of cases, some of which are discussed below, an equitable estoppel arises even though, at the time the assumption was induced and acted upon, the defendant was mistaken as to his or her own legal rights. Basing estoppel on carelessness in inducing the relevant assumption would require the courts in those cases to consider whether a reasonable person in the defendant's position would have inquired as to the true legal position. The courts in equitable estoppel cases do not ordinarily inquire as to these issues so it is not possible to say whether the defendants in estoppel cases behaved carelessly in inducing the relevant assumption. The notion of carelessness cannot therefore provide a compelling basis for understanding equitable estoppel. Accordingly, we cannot conclude that a defendant found liable on the basis of equitable estoppel committed a wrong in inducing the relevant assumption.

II. PROMISES AND RIGHTS-CREATING EVENTS

The analysis above shows that, while equitable estoppel cases often require the court to respond to a wrong, the wrong is not itself the rights-creating event.[46] Following Dixon J in *Grundt v Great Boulder Pty Gold Mines Ltd*, we may observe that a party who has relied upon an assumption induced by another has nothing to complain about '[s]o long as the assumption is adhered to.'[47] We could equally observe, however, that a promisee under a contract has nothing to complain about provided the promisor adheres to his or her contractual promise. In each case the statement obscures the nature and source of the rights that are contravened. What is of interest to scholars of contractual obligations, and should be of interest to scholars of reliance-based obligations, is not only

[45] Above n 23 [*Waltons*].

[46] As noted in S Waddams, *Dimensions of Private Law* (Cambridge, Cambridge University Press, 2003) 79.

[47] (1937) 59 CLR 641 (HCA) 674.

the breach of duty itself but also the series of events that create the obligation that has been breached. We need, in other words, to focus on the 'primary superstructure above the wrong'.[48] If equitable estoppel is not part of the law of wrongs, the taxonomic question that arises is whether it is concerned with rights arising by way of a promise, consent or some similar event which would bring it within the same broad category of obligation as contract.[49] This is a useful question to discuss because it requires us to consider important questions about the basis of the obligation in estoppel.

An obligation arises by way of estoppel from a series of events, consisting of (i) the inducement by A of an assumption which is adopted by B, and (ii) such action on the faith of that assumption by B that B would suffer detriment if A did not adhere to it. It is this chain of events (assumption, inducement, detrimental reliance) that creates rights in the relying party and a corresponding obligation in the person who induced the assumption. The making of a contract is a chain of events (offer and acceptance involving an exchange and manifesting an intent to be legally bound) which creates a legal entitlement to the benefits expected under the contractual agreement. As a statement of legal principle, we can say that the foundational chain of events in estoppel does not give rise to an entitlement to the expectation, but does give rise to an entitlement to protection from harm resulting from any inconsistent conduct by the defendant.[50] In contract, the failure to fulfil the promisee's expectation is a wrong. In estoppel, a wrong is committed when the defendant behaves inconsistently without taking steps to ensure that this does not cause harm to the relying party.[51] The defendant commits a wrong by infringing the claimant's right not to be harmed by the defendant's inconsistent conduct.

Having identified the rights-creating series of events in estoppel, we can now consider whether equitable estoppel is best seen as essentially contractual, or as a *sui generis* category of right and obligation which falls within the 'other causative events' category in the Birksian taxonomy of the law of obligations.[52] Stephen Smith has argued that 'the best overall explanation'

[48] Birks, above n 14, at 50.

[49] As to whether contract itself is best understood as a consent-based category of obligation, see A Robertson, 'On the Distinction between Contract and Tort' in A Robertson (ed), *The Law of Obligations: Connections and Boundaries* (London, UCL Press, 2004) 87; A Robertson, 'The Limits of Voluntariness in Contract' (2005) 29 *Melbourne University Law Review* 179.

[50] The principle is discussed at length in Robertson, above n 4. Peter Benson has advanced a justification for this difference between contract and estoppel: Benson, 'The Unity of Contract Law,' above n 41, especially at 175–6.

[51] *Cf* McFarlane, above n 31, at 159, who argues that, in the proprietary estoppel cases, no wrong is committed when the defendant attempts to act inconsistently because the claimant has already acquired 'a right based on the need to protect [the claimant's] reliance'.

[52] See, eg, Birks, above n 19, at 8.

of the English doctrine of promissory estoppel is that it is a promissory doctrine which can be seen as resting on the same foundations as contractual obligations.[53] Smith noted that this conclusion 'relies in part on moral arguments'.[54] A similar claim, though based more squarely on analysis of the case law, was made by Edward Yorio and Steve Thel to demonstrate that promissory estoppel in the United States is concerned with enforcing serious promises, rather than protecting reliance.[55] James Edelman has also argued that, although equitable estoppel is 'a creature that defies taxonomy', it is a doctrine which is essentially concerned with the enforcement of promises.[56] Although Smith was principally concerned with the English doctrine of promissory estoppel, the criteria he used provide a useful structure for assessing whether proprietary estoppel and the broader Australian doctrine are essentially reliance-based or promise-based doctrines. The three criteria are: first, whether a promise is required to establish liability; second, whether detrimental reliance is required to establish liability; and third, whether the remedies essentially involve the enforcement of promises.

A. Is a Promise Required?

A basic doctrinal difference between equitable estoppel and the principle of promissory estoppel described in section 90 of the *Restatement (Second) of the Law of Contracts* is that the former is based on an induced assumption whereas the latter requires a promise. It is unclear whether this is a difference of substance and whether it has any practical effect on the relative scope of the application of the doctrines.[57] The essence of a promise, both in ordinary speech[58] and in the *Restatement*,[59] is a commitment. In terms of accepted doctrine, it is clear that a promise is not

[53] Smith, above n 1, at 244. See also PS Atiyah and S Smith, *Atiyah's Introduction to the Law of Contract*, 6th edn (Oxford, Oxford University Press, 2005) 127.

[54] *Ibid*.

[55] E Yorio and S Thel, 'The Promissory Basis of Section 90' (1991) 101 *Yale Law Journal* 111. For a response see R Hillman, 'Questioning the "New Consensus" on Promissory Estoppel: An Empirical and Theoretical Study' (1998) 98 *Columbia Law Review* 580.

[56] J Edelman, 'Remedial Certainty or Remedial Discretion in Estoppel after Giumelli?' (1999) 15 *Journal of Contract Law* 179, 179. See also J Edelman, 'Taking Promises Seriously' (2007) 45 *Canadian Business Law Journal* 399.

[57] See Robertson, above n 20, at 44–5.

[58] The Oxford English Dictionary Online, <http://dictionary.oed.com> (last accessed: 2 November 2007) relevantly defines a promise as 'A declaration or assurance made to another person (usually with respect to the future), *stating a commitment* to give, do, or refrain from doing a specified thing or act, or *guaranteeing* that a specified thing will or will not happen' (emphasis added).

[59] The *Restatement (Second) of Contracts* (1981) § 2 defines a promise as 'a manifestation of intention to act or refrain from acting in a specified way, so made as to justify a promisee in understanding that a commitment has been made'.

required to found an equitable estoppel. In Australia, the threshold requirement is generally framed in terms of the defendant inducing an assumption as to the defendant's future conduct or the rights of the parties.[60] Similarly, the English Court of Appeal has accepted as 'the basic principles of proprietary estoppel' that a person can be prevented from asserting his or her legal rights if, short of a binding contract, the person makes a promise that he or she will not insist on them, and:

> Short of an actual promise, if he, by his words or conduct, so behaves as to lead another to believe that he will not insist on his strict legal rights—knowing or intending that the other will act on that belief—and he does so act, that again will raise an equity in favour of the other; and it is for a court of equity to say in what way the equity may be satisfied.[61]

Although the courts say that a promise is not required to found an equitable estoppel, one might nevertheless make the argument that a promise is required in practice. It might be argued, for example, that the only assumptions that can reasonably be relied upon are those that arise from promises.[62] Smith has questioned whether there is any conduct short of a promise on which one can justifiably rely. He suggests that 'a person who refrains from making a promise is thereby indicating that she is reserving the right to change her mind' and so a listener who relies on a non-promissory statement 'does so at his own risk'.[63] There is an obvious and well-recognised difficulty in distinguishing between conduct that amounts to a promise and conduct falling short of a promise, since a commitment to behave in a particular way in the future is usually implied rather than expressed. Nevertheless, one might question whether all conduct falling short of a promise necessarily communicates the reservation of a right to change one's mind.

Many proprietary estoppel cases arise from conduct that could properly be characterised as a promise, in the sense that the defendant has expressly or implicitly made a commitment to the claimant. Whether the defendant's conduct can be characterised as a promise is an important question in an estoppel case because reliance on a clear promise is more likely to be

[60] *Waltons*, above n 23, at 407, 428–9, 453, 458; *Commonwealth v Verwayen*, above n 23, at 412–13, 444, 460–61, 487, 500.

[61] *Jennings v Rice* (2002) [2003] 1 P & CR 8 (CA) [19] (Aldous LJ with whom Robert Walker LJ and Mantell LJ agreed), quoting *Crabb v Arun DC* [1976] Ch 179 (CA) 187 (Lord Denning MR).

[62] See S Smith, 'The Reliance Interest in Contract Damages and the Morality of Contract Law' [2001] *Issues in Legal Scholarship* 1, 23. *Cf* Benson, 'The Unity of Contract Law,' above n 41, at 174: 'First, there must be a promise or some other representation by one party upon which the other may reasonably rely. An essential condition of such reasonableness is that the second party must reasonably be able to construe the first party's intention as inviting reliance upon the promise or representation.'

[63] Atiyah and Smith, above n 53, at 127.

regarded as reasonable. A promise may even be 'perceived by reasonable persons as inviting reliance', as Benson has suggested.[64] A promise is, however, neither necessary nor sufficient to found a proprietary estoppel. Although many proprietary estoppel cases are founded on conduct that can accurately be characterised as a promise or commitment by the defendant, there is a second group of cases in which there is scope for argument about this. These are the cases in which the language of assurance or commitment is not expressly used, and it is arguable whether a commitment might have been implicit. There is, more significantly, a third group of cases in which proprietary estoppels arise in circumstances in which it is not possible to imply any promise or commitment. Proprietary estoppel can arise from 'acquiescence by one party in the known expectation of the other party that he has or will have a proprietary right or interest where the other party has acted to his detriment on that basis'.[65] The cases applying this principle clearly show that a promise is not required to establish a proprietary estoppel.[66]

We can begin with *Stiles v Cowper*, where a remainderman was prevented from denying the validity of an invalid lease created by a life tenant.[67] The remainderman accepted rent from the tenant, and stood by while the tenant expended £5,000 rebuilding the house. Lord Hardwicke LC said that while the acceptance of rent by itself would not bind the remainderman, 'when the remainder-man lies by, and suffers the lessee … to rebuild, and does not by his answer deny that he has notice of it, all of these circumstances together will bind him from controverting the lease afterwards'.[68] As Oliver LJ later noted, it appears from the report that the remainderman did not know at the time he accepted the rent and acquiesced in the expenditure that the lease was invalid. He accepted the rent because he 'thought proper' to do so.[69] In those circumstances, it is not possible to say that the landlord impliedly undertook to honour the lease. By accepting the rent and acquiescing in the expenditure, the landlord simply confirmed the tenant's assumption that the lease was binding. The landlord was bound by 'all these circumstances' not to act inconsistently with that assumption. Oliver LJ's influential decision in *Taylors Fashions Ltd v Liverpool Victoria Trustees Co Ltd* confirmed that an estoppel by acquiescence can arise even where the defendant was, at the

[64] Benson, 'The Expectation and Reliance Interests in Contract Theory,' above n 41, 61.

[65] *Scottish & Newcastle plc v Lancashire Mortgage Corporation Ltd*, above n 9, at [44].

[66] As Waddams, above n 46, at 60–61 has observed, citing *Ramsden v Dyson* (1866) LR 1 HL 129; *Spiro v Lintern* [1973] 3 All ER 319 (CA); *Crabb v Arun DC*, above n 61; *Waltons*, above n 23.

[67] (1748) 3 Atk 692, 26 ER 1198 (Ch) [*Stiles*].

[68] *Ibid*, at 1198.

[69] In *Taylors Fashions Ltd v Liverpool Victoria Trustees Co Ltd* [1982] QB 133 (ChD) 148.

time of the relevant conduct, mistaken as to his or her own legal rights.[70] Where this occurs, the concept of promise has no explanatory power, as the following recent cases show.[71]

In *Scottish & Newcastle plc v Lancashire Mortgage Corporation Ltd*, Lancashire Mortgage Corporation (LMC) advanced £30,000 to Pexman on the security, inter alia, of a charge over Pexman's house.[72] Pexman was at that time already indebted to Scottish & Newcastle (S&N) and, in fact, some of the money advanced by LMC went to reduce Pexman's indebtedness to S&N. S&N took a charge to secure the amount outstanding, and had its charge registered before LMC's. Pexman defaulted in repayment and the proceeds of the sale of the house were insufficient to meet both debts. A proprietary estoppel arose which prevented S&N from asserting the priority of its charge. The basis of the estoppel was that S&N knew that LMC expected that its charge would have priority over that of S&N, and stood by and allowed LMC to advance money on the faith of that expectation.[73] Mummery LJ expressly noted that this was 'a case of passive acquiescence by S&N rather than positive representation, encouragement or promise'.[74]

In *Munt v Beasley* a tenant improved a loft space on the mistaken assumption that the space was included in the lease.[75] The landlord was also mistaken as to the subject matter of the lease, but was aware that the building work was being undertaken and made no complaint about it. It was therefore held to be unconscionable for the landlord to assert his legal rights when he became aware of the true position. The Court of Appeal said that 'It would be unconscionable in this case, as Mr Beasley acquiesced in the works and Mr Munt suffered detriment in executing them in the belief that the loft was included in the Lease.'[76] In a case such as this there is no basis for saying that the defendant made a commitment or promise that his legal rights would not be enforced.[77] Indeed, in cases such as *Munt* and *Stiles*, where both parties are mistaken as to their legal rights, there cannot be an implied commitment or promise because the parties do not believe that there is anything to be making a commitment about.

The third recent English case is *Strover v Strover*.[78] The partners in an accounting firm effected 'mirror policies' of life insurance to protect one

[70] *Ibid*.

[71] See also *Roufeil v Lusby*, above n 9, where company directors who allowed the company to 'order its affairs' on the basis that it owned certain property were prevented from disputing that ownership.

[72] Above n 9.

[73] *Ibid*, at [47].

[74] *Ibid*. Sedley LJ and Moore-Bick LJ agreed.

[75] [2006] EWCA Civ 370 [*Munt*].

[76] *Ibid*, at [44].

[77] For another recent example see *Strover v Strover*, above n 10.

[78] Above n 10.

another in the event that one of them should die during the life of the partnership. The policies were intended to cover the expense that the surviving partners would incur in having to purchase the partnership interest of the deceased from his estate. Each policy was held on trust, with the two other partners nominated as beneficiaries. One of the partners died after he had retired from the partnership. At the time he retired he wrongly assumed that, following his retirement, the policy would enure for the benefit of his wife and children. In fact the other partners remained the beneficiaries under the trust. In reliance on that false assumption, the deceased lost the opportunity to renegotiate the 'mirror policies' arrangement before he died. The surviving partners would have been likely to agree to correct the arrangement if the matter had been raised with them because they were similarly exposed with their own policies. Hart J concluded that a proprietary estoppel arose in the circumstances, but held that the remedy should reflect the fact that there was a 20 per cent chance that the deceased would not have taken corrective steps had he not been mistaken. Accordingly, the equity arising by way of proprietary estoppel was held to be equal to 80 per cent of the proceeds of the policy. As in *Munt* and *Stiles*, since the defendants were mistaken as to their own legal rights, they cannot be said to have made any implicit promise or commitment.

It might be argued that, despite what the judges say, the basis of obligation in the 'acquiescence' cases involving assumptions of rights is different from that which underlies the 'promise' cases involving assumptions as to the future conduct of the promisor. There are three problems with this argument. The first is the difficulty of identifying an alternative basis of liability in the acquiescence cases. The obvious candidate is unjust enrichment, particularly in light of the Privy Council's confusion of the principles of unjust enrichment with those of proprietary estoppel in *Blue Haven Enterprises Ltd v Tully*.[79] Birks once argued that the courts should 'distinguish in this context between the cause of action based on promises and the cause of action based on free acceptance'.[80] He said that 'Where there is no promise or expectation-inducing conduct, only the cause of action in free acceptance should be recognised and only the restitutionary measure of recovery should be allowed.'[81] In improvement cases such as

[79] [2006] UKPC 17. A claim 'based on unjust enrichment' relating to the improvement of land was rejected on the ground that the defendant did not acquiesce in the improvement or make any relevant representation to the claimant. The case is discussed by McFarlane, above n 31 and by K Low, 'Unjust Enrichment and Proprietary Estoppel: Two Sides of the Same Coin?' [2007] *Lloyd's Maritime and Commercial Law Quarterly* 14.

[80] P Birks, *An Introduction to the Law of Restitution*, rev edn (Oxford, Clarendon Press, 1989) 293.

[81] *Ibid*. See further N Hopkins, 'Estoppel and Restitution: Drawing a Divide' in E Cooke (ed), *Modern Studies in Property Law* (Oxford, Hart Publishing, 2003) vol 2, 145.

Stiles, it is possible to point to an enrichment which the defendant has received at the expense of the claimant.[82] There are, however, acquiescence cases such as *Strover v Strover* in which the defendants were not in any relevant sense enriched by the actions of the claimant. Moreover, where the acts of reliance do enrich the defendant, the remedies granted by way of estoppel redress the detriment resulting from reliance, rather than restoring the enrichment.[83] In *Munt*, for example, the court had before it evidence as to the amount by which the tenant's DIY loft-conversion increased the value of the defendant's land,[84] but ignored this in favour of a remedy which was 'proportionate to the expenditure of money and time on the conversion'.[85] In *Scottish & Newcastle plc v Lancashire Mortgage Corporation Ltd*, the defendant was clearly enriched, since some of the money advanced by LMC (£20,000) was used to reduce the debtor's liability to S&N.[86] The effect of the estoppel, however, exceeded the amount of the enrichment (£20,000), since it gave LMC priority for the entire amount advanced (£30,000). The estoppel operated to the extent of the claimant's reliance interest, which happened in this case to coincide with its expectation interest.

The second problem with attributing the cases involving assumptions of rights to a different equity from those involving assumptions as to future conduct is that in some fact situations it is difficult to distinguish between them. In *Waltons*,[87] for example, where negotiations for a lease were concluded and the documents had been signed by the landlord-claimants and sent to the tenant-defendant for signature, the relevant assumption could be characterised in different ways. It could be characterised as an assumption that the tenant would sign the lease (an assumption as to future conduct), that the tenant had signed the lease (an assumption as to existing fact) or that a binding agreement had come into effect between the parties (an assumption as to rights). It would be most unsatisfactory if significant consequences flowed from the fine distinction between assumptions as to rights and assumptions as to future conduct, in addition to those which already flow from the fine distinction between assumptions of fact (raising the evidentiary doctrine of common law estoppel by representation) and assumptions as to rights or future conduct (raising the substantive doctrines of equitable estoppel).

[82] As Low, above n 79, at 16, has noted, however, the courts in the acquiescence cases tend to ignore the presence or absence of an enrichment.

[83] *Cf* Waddams, above n 46, at 59, noting that enrichment is not always present and does not normally shape the remedy, but suggesting that the prevention of unjust enrichment is not irrelevant to proprietary estoppel.

[84] Above n 75, at [5].

[85] *Ibid*, at [45].

[86] Above n 9.

[87] Above n 23.

The third problem with drawing a distinction between assumptions of rights and assumptions as to future conduct is that assumptions as to future conduct can also arise by way of acquiescence. *Waltons* could be seen as an example of this. A more recent example is *Hypec Electronics Pty Ltd v Mead*.[88] Mr and Mrs Mead, the directors of Hypec, used the company's money to purchase certain properties. It was not clear whether the money had been lent to the directors or the directors held the properties on trust for Hypec. The Meads' marriage broke down and Hypec went into liquidation. The liquidator initially treated the property transactions as loans made to the directors by Hypec. Mr Mead sought to use the properties to fund expensive litigation disputing a debt for which judgment had been entered against Hypec in favour of a company associated with Mrs Mead. The liquidator allowed Mr Mead to incur considerable costs proceeding with the litigation on the assumption that the properties would be available to fund the litigation. By allowing Mr Mead to obtain orders for the sale of the properties in Family Court proceedings between the directors, and by failing to reserve his right to assert Hypec's interest in the properties, the liquidator was held to have induced the assumption that he would allow the properties to be used in this way and would not assert the company's interest. The liquidator was therefore held to be estopped from asserting Hypec's entitlement to the properties.[89] The relevant assumption was induced by acquiescence, but it clearly related to the liquidator's future conduct rather than the director's existing rights in the properties.[90] It would be extremely artificial to say in these circumstances that the liquidator made a promise or commitment to act in a particular way. The estoppel arose because the liquidator had led Mr Mead to believe that he would behave in a particular way, namely that he would allow the properties to be used to fund the litigation and not assert the company's interest in them, and he knew that Mr Mead was incurring substantial legal costs on the faith of that assumption.

Although the conduct on the part of the defendant that gives rise to a proprietary estoppel can in many cases accurately be characterised as a commitment or a promise, the doctrine cannot be characterised as a promissory doctrine because there is a significant minority of cases in which estoppels arise from conduct that clearly falls short of a promise. Nor can we conclude that the obligation in estoppel cases is voluntary or consensual. Michael Pratt has suggested that the orthodox position is that

[88] Above n 9.

[89] An order was made to ensure that the remedy was limited to Mr Mead's reliance loss. Relief was conditional on Mr Mead undertaking to inform the liquidator of the progress of attempts to recover his costs from the other parties to the proceedings, and the liquidator was given liberty to apply for an order for repayment of any amounts so recovered: *ibid*, at [199].

[90] The assumption on which the estoppel was founded was that 'the liquidator would not stand in the way of the orders being implemented': above n 9 (CA), at [46].

'an obligation is voluntary if and only if it depends for its validity on the intention of the obligor to acquire it, which intention is a positive reason for its existence'.[91] There are three reasons why the obligation in proprietary estoppel cannot be regarded as voluntary or consensual, even on this relatively undemanding definition.[92] First, unlike contract, the principles of estoppel do not require that the defendant manifest an intention to assume a legal obligation. Second, in cases like *Munt*, where the defendant believes the claimant is entitled to do what he or she is doing, the defendant clearly cannot intend to assume an obligation. Third, even in those cases in which the defendant makes an express promise or commitment to behave in a particular way in the future, the obligation that arises from estoppel is not coextensive with the commitment or promise. Proprietary estoppel does not give rise to an obligation to do what the defendant has promised to do but, as we will see shortly, simply gives rise to an obligation not to cause harm by behaving inconsistently.

B. Is Detrimental Reliance Required?

A crucial part of Yorio and Thel's argument that the basis of promissory estoppel under section 90 of the *Restatement* is promise, rather than reliance, was that the elements of inducement and detriment are neither necessary nor sufficient to establish liability.[93] Similarly, Smith has questioned whether detrimental reliance is a condition of liability in promissory estoppel in England. There is good reason to doubt the role of reliance in that doctrine, since detriment is not consistently required in promissory estoppel cases.[94] Smith's scepticism about the detrimental reliance requirement, however, seems to extend to the substantive doctrines of estoppel:

> What is needed, in order to prove the reliance view [of promissory estoppel], is evidence of contract cases in which the courts denied estoppel arguments on the basis of the reliance requirement. But it is not clear that any such cases exist. …In jurisdictions in which estoppel can be used as a cause of action, it should, in theory, be easier to find cases in which estoppel claims are denied because the plaintiff who should rely, did not …[95]

The example Smith seeks is of a situation in which a contractual obligation is waived by the plaintiff and not performed by the defendant, but in which 'it was proven that the defendant would not have performed regardless of

[91] M Pratt, 'Promises, Contracts and Voluntary Obligations' (2007) 26 *Law and Philosophy* 531, 535.

[92] *Cf* Robertson, 'The Limits of Voluntariness in Contract,' above n 49.

[93] Yorio and Thel, above n 55, at 151–61.

[94] See E Cooke, *The Modern Law of Estoppel* (Oxford, Oxford University Press, 2000) 100–103; Neyers, above n 8, at 31–3.

[95] Smith, above n 1, 237–8 (footnote omitted).

the plaintiff's waiver—in other words, that the plaintiff [this should be the defendant] did not rely on the waiver'.[96] In *Brikom Investments Ltd v Carr*[97] it was argued unsuccessfully that the estoppel claim should fail because one of the estoppel claimants would have acted as she did in the absence of the representation. As Lord Denning MR explained, the courts are unwilling to speculate about this counterfactual question where a party enters into a transaction following the making of a promise or representation by another:

> It is no answer for the maker to say: 'You would have gone on with the transaction anyway.' That must be mere speculation. No one can be sure what he would, or would not, have done in a hypothetical state of affairs which never took place … Once it is shown that a representation was calculated to influence the judgment of a reasonable man, the presumption is that he was so influenced.[98]

Reliance in this context means a sufficient causal connection between the assumption induced by the defendant and the detrimental change of position by the claimant. The law gives significant evidentiary advantages to an estoppel claimant in relation to proof of this causal connection. Where the claimant has been induced to adopt an assumption and subsequently acts to his or her detriment, a rebuttable presumption of reliance arises.[99] Moreover, reliance on the relevant assumption need not be the claimant's only reason for taking the detrimental action: it is enough if it was a reason for acting.[100] The same approach is used for causation in relation to duress and misrepresentation.[101] The point is not that reliance is unimportant but that whether a person has relied is a question relating to that person's state of mind, which cannot easily be proved or disproved. If detrimental action has been taken following the inducement of an assumption then an attempt to identify the role played by each of the different factors that may have contributed to the decision to act is not likely to be illuminating. Lord Justice Cranworth put the point even more strongly in *Reynell v Sprye*, where he said that

> It is impossible so to analyze the operations of the human mind as to be able to say how far any particular representation may have led to the formation of any

[96] *Ibid*, at 237.
[97] [1979] QB 467 (CA).
[98] *Ibid*, at 482–3.
[99] *Campbell v Griffin* [2001] EWCA Civ 990, [25]; *Wayling v Jones* (1993) 69 P & CR 170 (CA); *Greasley v Cooke* [1980] 3 All ER 710 (CA) 713; *Coombes v Smith* [1986] 1 WLR 808, 821.
[100] *Campbell v Griffin*, *ibid*; *Wayling v Jones*, *ibid*, at 173.
[101] See, eg, A Robertson, 'Partial Rescission, Causation and Benefit' (2001) 17 *Journal of Contract Law* 163, 168–73; R Bigwood, *Exploitative Contracts* (Oxford, Oxford University Press, 2003) 347–51.

> particular resolution, or the adoption of any particular line of conduct. No one can do this with certainty, even as to himself, still less as to another.[102]

The decision of the English Court of Appeal in *Campbell v Griffin* shows how difficult it is to disentangle reliance from other motives.[103] The claimant was a lodger who developed a close relationship with his landlords. He undertook domestic work and provided personal care for them, having been repeatedly assured that he had a home for life. The claimant said in his witness statement that he would have moved out had it not been for the assurances, but admitted under cross-examination that he was also influenced by his friendship with the landlords and a concern for their welfare. The trial judge found against the claimant on the basis that the detrimental action taken by the claimant had not been caused by his reliance on the assurances made by the defendants. The decision was overturned by the Court of Appeal on the ground that the trial judge had given insufficient weight to the presumption of reliance and had attached too much importance to the claimant's concessions about his mixed motives. He had done much more than could be expected of even the most friendly lodger. The court held:

> In cases of this sort it is inevitable that claimants should be asked hypothetical questions of the 'what if' variety but the court is not bound to attach great importance to the answers to such hypothetical questions. …it would do no credit to the law if an honest witness who admitted that he had mixed motives were to fail in a claim which might have succeeded if supported by less candid evidence.[104]

It is clear from the above discussion that the courts will usually lack the factual foundation necessary to scrutinise closely the issue of causation in estoppel cases. There can, however, be no doubt that detrimental reliance is a fundamental requirement of both proprietary estoppel[105] and the broader Australian doctrine.[106] A search of the LexisNexis databases confirms that the English and Australian courts routinely deny claims of proprietary (and equitable) estoppel on the basis that the claimant would not suffer material detriment as a result of his or her reliance if the defendant was allowed to act inconsistently with the relevant assumption. In many cases this is given as a reason, amongst others, for denying liability.[107] There are, however,

[102] (1852) 1 De GM & G 660, 42 ER 710 (Ch) 728.

[103] Above n 99.

[104] *Ibid*, at [28]–[29].

[105] See, eg, *Jennings v Rice*, above n 61, at [21]: 'There can be no doubt that reliance and detriment are two of the requirements of proprietary estoppel'; *Gillett v Holt* [2001] Ch 210 (CA) 232: 'The overwhelming weight of authority shows that detriment is required.'

[106] *Waltons*, above n 23, at 404, 429; *Commonwealth v Verwayen*, above n 23, at 413, 429, 444, 455, 500. See Robertson, above n 20, at 44.

[107] For example *Gordan v Mitchell* [2007] EWHC 1854 (Ch); *Armitage Holdings Inc v Delahunty* [2007] EWHC 1556 (Ch); *Wakelam v Boardman* [2007] NSWSC 567;

clear instances of cases in which the lack of detrimental reliance was the only reason for rejecting the estoppel claim.[108] In light of cases such as *Jennings v Rice*[109] and *Sledmore v Dalby*,[110] it cannot be doubted that detrimental reliance is a crucial element of proprietary estoppel. The requirement of detriment is also strictly applied in cases involving the broader Australian doctrine of equitable estoppel. It is not difficult to find examples of cases in which, even though a promise was made and acted upon by the claimant, the equitable estoppel claim failed because the claimant's action was held not to have been detrimental. A recent example is *St George Soccer Football Association Inc v Soccer NSW Ltd*.[111] The claimants were soccer clubs excluded from the defendant's 'premier league' competition when it was reduced in size from 16 clubs to 10. The defendant had earlier told the claimants and other clubs that applicants for the new competition were required to meet certain specified criteria.[112] The defendant later changed its policy and did not adhere to the criteria in selecting the clubs that would participate. The court accepted that the representation was made and acted upon by the claimants in submitting their applications to join the new competition. The court continued:

> But it cannot be said that they thereby acted to their detriment. The alternative course of action would have been to refrain from submitting applications. Had they done that, their ultimate position would have been the same as that they in fact came to occupy, namely, non-participation in the new competition. The requirements for the creation of an estoppel are therefore not established.[113]

There can be no doubt that detrimental reliance is an essential condition of liability for both proprietary estoppel and the broader Australian doctrine.

Comptroller-General of Customs v Parker (2006) 200 FLR 44 (NSWSC); *Yarmouth Harbour Commissioners v Harold Hayles Ltd* [2004] EWHC 3375 (Ch); *Bredel v Moore Business Systems Australia Ltd* [2003] NSWCA 117; *Hawker Pacific Pty Ltd v Helicopter Charter Pty Ltd* (1991) 22 NSWLR 298. See also *Rosenberg v Fifteenth Eestin Nominees Pty Ltd* [2007] VSC 101 and *Mobil Oil Australia Ltd v Wellcome International Pty Ltd* (1998) 81 FCR 475 where the claims failed because any detriment suffered as a result of reliance by the claimants was not proportional to the relief sought.

[108] For example *Hunt v Soady* [2007] EWCA Civ 366; *Fisher v Brooker*, above n 11; *St George Soccer Football Association Inc v Soccer NSW Ltd* [2005] NSWSC 1288; *Etchison v ANZ Executors and Trustee Company Ltd* [2005] QSC 363; *Sea Food International Pty Ltd v Theng Pew Lam*, unreported, Federal Court of Australia, Cooper J, 27 February 1998, BC9800437.

[109] Above n 61.

[110] Above n 23.

[111] Above n 108. Another example is *Sea Food International Pty Ltd v Theng Pew Lam,* above n 108.

[112] Above n 108, at [141].

[113] *Ibid*, at [175].

C. Are the Remedies Promissory?

The third argument advanced in favour of the promise-based view of estoppel is that the remedies typically involve the enforcement of the promise.[114] Smith asks:

> Do these remedies protect the complainant's 'promissory' interest in having a promise performed (as the promissory view asserts) or do they protect the claimant's 'reliance' interest in not suffering a reliance-based loss (as the reliance view asserts)? Here again, the evidence does not give a clear answer. In English law, which generally does not permit estoppel to be used as a cause of action, the issue of the appropriate measure of damages for an estoppel does not arise.[115]

Even in the 'exceptional' cases in which estoppel is permitted to operate as a cause of action, Smith suggests, the approach to relief supports the promissory view of estoppel:

> Under the doctrine of proprietary estoppel, for example, a landowner who represents to another that the owner's land will be conveyed to him will be required to do so if the other relies on that representation—even if the value of the reliance is less than the value of the land. The remedy in such cases is always specific performance—which, of course, is entirely consistent with promissory conceptions of remedies.[116]

The recent English and Australian cases clearly contradict the promissory conception of remedies.[117] Both English and Australian law now require the relief granted in a proprietary estoppel case to be proportionate to the detriment that has been or would be suffered by the claimant. In *Jennings v Rice*, Aldous LJ described this as 'the most essential requirement' in the determination of relief.[118] The Australian and English courts have given explicit consideration to that requirement in the great majority of the recent cases.[119] Although the courts are careful to go no further than is necessary to protect the claimant's reliance interest, in most cases the only way to do so is by granting expectation relief, either *in specie* or in monetary form. The principal reason for this is that the detriment that is suffered in proprietary estoppel cases tends to be of a type that is difficult to quantify. Common examples involve the performance of services that the claimant is not in the business of performing, emotional investment in a

[114] In relation to s 90, see again Yorio and Thel, above n 55.
[115] Smith, above n 1, at 238.
[116] Smith, above n 1, at 238–9.
[117] This and the following paragraph draw on the analysis of the recent cases in Robertson, above n 4. See further A Robertson, 'Satisfying the Minimum Equity: Equitable Estoppel Remedies after *Verwayen*' (1996) 20 *Melbourne University Law Review* 805; A Robertson, 'Reliance and Expectation in Estoppel Remedies' (1998) 18 *Legal Studies* 360.
[118] Above n 61, at [36].
[119] See Robertson, above n 4.

family home developed through long-term residence, lost career opportunities and other life-changing decisions such as having children. Lost commercial ventures and investment opportunities are also difficult to quantify.[120] The courts cannot accurately compensate such detriment, but can prevent it by fulfilling the claimant's expectations. The difficulty that the courts experience in quantifying reliance loss in most proprietary estoppel cases justifies approaching relief on the basis that the claimant has a prima facie right to a remedy in the expectation measure. It also justifies the principle that the remedy should be proportional to the claimant's reliance loss, rather than having a rule that the remedy should precisely correspond to the reliance loss.

In a significant minority of the recent proprietary estoppel cases both the English and Australian courts have refused to grant relief in the expectation measure on the basis that such relief would be disproportionate to the claimant's reliance loss.[121] In some cases the courts have adopted a mathematical approach, awarding the claimant compensation calculated precisely by reference to reliance loss.[122] In *Young v Lalic*, for example, the claimant contributed $50,000 to the construction of a house in the expectation of a half-interest in property worth $800,000. Since she suffered no other detriment, Brereton J held that fulfilment of her expectations would be a disproportionate response, and her equity was held to be satisfied by a charge for her contribution, together with interest. In cases where the claimant's reliance loss could not accurately be quantified, the courts have made a 'broad brush' determination of an amount of compensation that is proportional to the detriment suffered by the claimant.[123] In one recent Australian case, proprietary estoppel had the effect of suspending the defendants' rights in order to give the claimant an opportunity to resume her original position. In *Sullivan v Sullivan* the claimant had given up subsidised public accommodation for which she waited seven years on the faith of a promise of a home for life.[124] The effect of the estoppel was to allow the claimant to remain in the promised house for a further seven years in order to give her an opportunity to return to public housing.

The routine awarding of expectation-based relief in estoppel cases can comfortably be reconciled with the view that the doctrine is concerned to protect against harm resulting from reliance rather than to fulfil promises.

[120] For example *Van Laethem v Brooker* [2005] 2 FLR 495 (Ch).

[121] See, eg, *Jennings v Rice*, above n 61; *Powell v Benney* (2007) [2008] 1 P & CR DG12 (CA).

[122] *Young v Lalic*, above n 37; *Repatriation Commission v Tsourounakis* (2007) 158 FCR 214; *Strover v Strover*, above n 10. See also the reliance-based condition attached to the remedy granted in *Hypec Electronics Pty Ltd v Mead*, above n 9, mentioned in n 89 above.

[123] *Jennings v Rice*, above n 61; *Ottey v Grundy* [2003] EWCA Civ 1176; *Donis v Donis* [2007] VSCA 89.

[124] [2006] NSWCA 312.

In most cases there is no other way to ensure protection against harm, because the detriment cannot accurately be quantified. There is not, however, a similar promise-based explanation for the significant minority of cases in which the courts grant more limited relief. We must therefore conclude that the remedial goal is to protect against harm.

III. CONCLUSION

Although equitable estoppel routinely responds to conduct that could be characterised as a wrong or apprehended wrong, equitable estoppel is not properly seen as part of the law of wrongs. That is because in equitable estoppel, as in contract, it is possible to identify a series of events which give rise to primary rights that are recognised by the law prior to and independently of any infringement of those rights. This significant feature of equitable estoppel distinguishes it from tort. In an events-based classificatory scheme, equitable estoppel is *sui generis*: the rights-creating event is detrimental reliance on an assumption induced by another. To describe proprietary or equitable estoppel as 'reliance-based,' however, begs the question as to the basis of the obligation. As Stevens has observed, reliance alone is clearly an insufficient basis for liability.[125] Although detrimental action in reliance is undoubtedly necessary for liability in equitable estoppel, and constitutes the final step in the rights-creating series of events, it is ultimately that which the claimant has relied upon that justifies the defendant's liability. It is clear from the above discussion that it is not the fact the defendant has made a promise that justifies his or her liability. A promise is neither necessary nor sufficient to establish liability. Nor can we say that the defendant is responsible for having behaved carelessly in inducing reliance. The basis of equitable estoppel is an obligation not to cause harm through inconsistent conduct. Where A has played a role in inducing the adoption of an assumption by B, and A knows, intends, or should reasonably expect that B may rely on the assumption,[126] then those circumstances give rise to a duty on the part of A not to cause harm to B by behaving inconsistently with the assumption.

[125] Stevens, above n 17, at 15.

[126] See A Robertson, 'Knowledge and Unconscionability in a Unified Estoppel' (1998) 24 *Monash University Law Review* 115.

9

Lumley v Gye *and the (Over?)Protection of Contracts*

G H L FRIDMAN*

LUMLEY v GYE belongs in that group of nineteenth-century cases that helped to form the fundamental principles of our modern common law.[1] Notwithstanding the undeniable importance of the case, the impression I have is that on the whole more attention has been paid, especially by academic writers, to the effect of the decision than to the purpose behind it and the language in which it was expressed.[2] That language merits analysis, not least because the case is an illustration of 'judicial activism'. I mean by this phrase, which is so frequently employed in modern times particularly, but not exclusively, in relation to the judgments of the Supreme Court of Canada, the fashioning by a court of decisions designed to achieve certain purposes thought desirable by the court, regardless of the correctness of those decisions in terms of the doctrine of precedent.

Recently Lord Hoffmann declared that *Lumley v Gye* gave rise to what he termed 'accessory liability', a new type of tortious responsibility.[3] In effect, he was suggesting the liability recognised in the case emerged from the judges in the majority of the court much as the goddess Athene sprang

* I am greatly indebted to Stephen Pitel whose trenchant criticism of earlier drafts preserved me from many errors of taste, clarity and purpose. For those, if any, that remain, I am solely responsible.

[1] (1853) 2 El & Bl 216, 118 ER 749 (QB).

[2] FB Sayre, 'Inducing Breach of Contract' (1923) 36 *Harvard Law Review* 663; CE Carpenter, 'Interference with Contract' (1928) 41 *Harvard Law Review* 728; Note, 'Tortious Interference with Contractual Relations in the Nineteenth Century: The Transformation of Property, Contract, Tort' (1980) 93 *Harvard Law Review* 1510; D Dobbs, 'Tortious Interference with Contractual Relationships' (1980) 34 *Arkansas Law Review* 335; H Perlman, 'Interference with Contract and Other Economic Expectancies: A Clash of Tort and Contract Doctrine' (1982) 49 *University of Chicago Law Review* 61; R Bagshaw, 'Can the Economic Torts be Unified?' (1998) 18 *Oxford Journal of Legal Studies* 729; AP Simester and W Chan, 'Inducing Breach of Contract: One Tort or Two' (2004) 63 *CLJ* 132.

[3] *OBG Ltd v Allan* (2007) [2008] 1 AC 1 (HL) [8]. *Cf* Lord Nicholls, *ibid*, at [172].

from the head of Zeus. As such it was a prime example, though of course not the only one in nineteenth-century England, of judicial activism. Whenever this occurs it is pertinent to inquire whether what the particular court did was justified in purely legal terms and, in so far as this is deducible, what non-legal factors led the court to decide as it did.

I. THE DECISION ANALYSED

The issue raised by the demurrer pleaded by the defendant was whether Lumley was entitled to sue Gye for persuading Johanna Wagner to abandon her contract with Lumley to sing in his theatre so that she could instead sing for Gye in his. It is an indication of its novelty and difficulty that four months passed before four of the five judges of the Court of Queen's Bench (the Chief Justice, Lord Campbell, not participating) reconvened to give judgment. The majority in favour of allowing the action to proceed consisted of Crompton, Erle and Wightman JJ. Coleridge J dissented.

That what was involved in this case was not cut and dried had become clear when counsel addressed the court. In 1853 there unquestionably existed an action against someone who enticed a servant away from his master.[4] Its origin was disputed. Crompton and Wightman JJ considered that it arose at common law. Coleridge J embraced the view that the cases which allowed the action derived from the provisions of the Statute of Labourers of 1351, which supplemented the Ordinance of Labourers of 1349.[5] Both were passed because of the labour shortage resulting from the Black Death.[6] Coleridge J established, partly by reference to Year Book cases, that the action came into being following the criminal liability set out in the statute.[7] Even earlier, however, there was authority that supported liability where a master was deprived of his servant's services because of a beating administered by the defendant. The basis of this liability was the violence employed.[8] Only after 1351 did there emerge liability for such interference without the use of violence, merely by seducing the servant away from the original employer.

The proposition that at common law an action lay at the suit of a master against one who caused the breach of the master's contract with his servant, either by harming or killing the servant or by enticing him away, seems to conflict with the fundamental doctrine of privity of contract. In *Bowen v Hall* Brett LJ clearly indicated that, as regards the origins of the

[4] Sayre, above n 2, at 663–72.
[5] 25 Edw III and 23 Edw III respectively.
[6] Dobbs, above n 2, at 340 note 27.
[7] *Cf* Sayre, above n 2, at 663–72.
[8] Dobbs, above n 2, at 338–40.

action for enticing a servant away from his master, the view of Coleridge J in *Lumley v Gye*, explained above, was preferable to that of the majority.[9] Those narrow origins did not support the broader decision in *Lumley v Gye*. The true ratio decidendi of that case, the basis on which it could be supported and endorsed, was said in *Bowen v Hall*, much more broadly, to be that:

> wherever a man does an act which in law and in fact is a wrongful act, and such an act as may, as a natural and probable consequence of it, produce injury to another, and which in the particular case does produce such an injury, an action on the case will lie.[10]

This was stated to have been derived from the much earlier case of *Ashby v White*.[11] However that decision was not cited, nor is this ratio to be found, in the judgments of Crompton and Erle JJ in *Lumley v Gye*. A version of this ratio, but without attribution to *Ashby v White*, does appear in the judgment of Wightman J as one, but not the only nor the governing, reason for permitting the action. Wightman J seems to concur with, and rely on, the ratio in the judgment of Crompton J, namely that, by what Crompton J called 'strict analogy' with the master–servant situation referred to earlier, the same action could be brought where the contract in issue was not a contract of service.[12]

Erle J gave an entirely different reason for allowing Lumley's action to proceed. He responded to the objection that, since actions for breach of a contract of hiring were based on no principle, such actions could not be extended beyond existing precedents, which related to contracts respecting trade, manufactures or household service but not performance at a theatre. His response was that actions for procuring the breach of a contract of hiring were indeed based on a principle. That principle, he said, was that the injured party had a right to the service and 'the procurement of the violation of the right is a cause of action'.[13] From this he proceeded to the more general notion that the violation of a right was a cause of action where the violation was an actionable wrong, as in a violation of a right to property or personal security. From this he concluded that 'when this principle is applied to a violation of a right arising upon a contract of hiring, the nature of the service contracted for is immaterial'.[14] Therefore it did not matter that Joanna Wagner was not a servant in the strict legal sense.

[9] (1881) 6 QBD 333, 339–40.
[10] *Ibid*, at 337.
[11] (1703) 1 Holt KB 524, 90 ER 1188.
[12] *Lumley v Gye*, above n 1, at 229.
[13] *Ibid*, at 232. *Cf Mogul SS Co Ltd v McGregor Gow & Sons* (1889) 23 QBD 598, 614 (Bowen LJ).
[14] *Ibid*.

Although the wrongfulness of the act was stressed in *Lumley v Gye* as being the basis of liability, 'malicious behaviour', the nature of which was undefined and unexplained, was also mentioned as possibly relevant. Yet the court did not countenance the view that 'malice' made actionable that which otherwise was not actionable. Such an effect of malice was negated, in *Bowen v Hall*,[15] by Coleridge J's son, John Duke Lord Coleridge.[16] In that case the majority upheld and approved *Lumley v Gye*, over the dissent of Lord Coleridge who, like his father before him, disagreed with the decision. In arriving at this conclusion Brett LJ relied on the wrongfulness in law and fact of what was done rather than on malicious behaviour.[17] However, Brett LJ, by then Lord Esher MR, brought malice back into the picture when, in his dissenting judgment in *Mogul SS Co Ltd v McGregor Gow & Sons*, he suggested that liability under *Lumley v Gye* involved acting with the 'malicious intention' of interfering with the plaintiff's right.[18] By malice he meant 'a malicious motive'. Later, in *Temperton v Russell* it was clearly stated that the *Lumley v Gye* tort could be committed even though the relevant contract was not one under which personal service was provided by one party to the other.[19] According to Lord Esher, inducing or procuring the breach of a contract was wrongful because it was done 'maliciously'.[20] The other members of the court, AL Smith and Lopes LJJ, did not use the same language but the general tenor of their judgments was to the same effect.[21]

The net result of these three cases was to create a new tort that girded contracts with a protective shield similar to, but more real than, the magic fire surrounding Brunnhilde. What the courts did was to impose liability for acts that were held to be impermissible, in the words of one American commentator, 'only because the defendant's purpose is thought insufficiently laudable'.[22] Why was this? One reason sometimes offered, and perhaps more often just assumed, says the same commentator, 'is that interference with a contract should produce liability because it is wrong to interfere. This is, however, very much the same as saying that it is wrong because it is wrong.'[23] In other words, the conclusion reached by the courts was foregone. It was dictated by the very nature of the question, given the context. My own view is that these criticisms have considerable force.

[15] Above n 9, at 342–3.
[16] The first Lord Chief Justice of England: W Holdsworth, *History of English Law* (London, Methuen & Co Ltd, 1965) vol 15, 460–66.
[17] *Bowen v Hall*, above n 9, at 338.
[18] *Mogul SS Co Ltd v McGregor Gow & Sons*, above n 13, at 608–609.
[19] [1893] 1 QB 715.
[20] *Ibid*, at 728.
[21] *Ibid*, at 730, 732–3.
[22] Dobbs, above n 2, at 343.
[23] *Ibid*.

What may have contributed to this idea that interference with a contract was wrong in itself is the frequently repeated notion of 'the sanctity of contract'. Such language seems to elevate the mundane, though useful, concept of contract into something more sublime. As a recent English commentator remarked, 'the phrase is all very well, but contract law is largely about commerce, not holiness'.[24] He underlined the point further by saying: 'Contract law is a public service the state offers to people who want to use it—rather like the National Health Service.'[25]

What *Lumley v Gye* did, said this author, was to turn into a tort the everyday commercial act of offering someone a better deal, enshrining a rule that was inimical to competition.[26] To which it might be added that competition could be considered an integral part of a free society, so that anything thought to hinder competition, such as the doctrine emanating from *Lumley v Gye*, might be undesirable. Hence, perhaps, the suggestion by Sayre in 1923 that 'It will not do to make a fetish of this tort remedy for the better protection of contractual rights.'[27]

Lumley v Gye made a profound change in the law. It gave birth to the tort of procuring or inducing a breach of contract. This, in turn, spawned the tort of wrongful interference with trade or business by the use of unlawful means. Beyond these developments in the law of torts, in creating accessory liability for breach of contract the decision also had consequences for the concepts of contract, contractual rights and privity. The subsequent development of the *Lumley v Gye* tort, I suggest, brings to the fore the clash between morality and legal rights.

Before delving into these matters, however, one intriguing aspect of the case merits mention. The action was permitted to proceed as a result of the decision of the Court of Queen's Bench. At the trial before a special jury, the plaintiff Lumley lost because, as Waddams has explained, he could not establish the malicious conduct required on the part of the defendant for there to be liability.[28] This, in turn, was because the defendant believed, or at any rate convinced the jury that he believed, that the contract between the plaintiff and Wagner had ceased to exist or had been terminated legally by Wagner because the plaintiff had not paid her an agreed advance. This trial was presided over by Lord Campbell, the Chief Justice of the Court of Queen's Bench. As previously noted, he did not participate in the earlier proceedings arising from the demurrer. One might consider his absence

[24] D Howarth, 'Against *Lumley v Gye*' (2005) 68 *MLR* 195, 202.
[25] *Ibid*, at 203.
[26] *Ibid*, at 202.
[27] Sayre, above n 2, at 686.
[28] S Waddams, *Dimensions of Private Law* (Cambridge, Cambridge University Press, 2003) 36–8.

from the court on that occasion, given the importance and novelty of the issue at stake, to be somewhat odd. Speculation as to his reasons, however, is idle.

If the role of Lord Campbell in this litigation is something of a mystery, what was done by Crompton, Erle and Wightman JJ is extremely clear. They chose to enlarge the scope of earlier decisions, to ignore the historical relevance of violence or criminal activity contrary to the 1351 statute as forming the basis of liability, to consider that they were not confined by the language of that statute, and to introduce the idea of malicious behaviour and the concept of violation of rights as grounds for creating liability. Their decision was revolutionary. There was no pragmatic reason or justification for their decision. The Wagner situation did not give rise to the possibility that she would not be able to compensate Lumley—one reason given by Crompton J for allowing the action, with which can be contrasted the suggestion in a difficult and confusing passage in the judgment of Erle J that the only recourse for the innocent victim of a contract procured to be broken is to sue the contract-breaker, not the procurer of the breach.[29] Lumley could have sued Wagner for breach of contract and been compensated by damages.

In fact, as the court also knew, Lumley had successfully sought an injunction against Wagner to prevent her from appearing for Gye, the effect of which would be to force her to appear for Lumley or not appear at all. Why then did he also wish to sue Gye? Could he have obtained different damages from those recoverable from Wagner, as Crompton J suggested?[30] Was the real reason for the action the vendetta or rivalry between these two theatre managers? If so, why did the Court of Queen's Bench consider its judicial duty was to create a remedy that would enable one rival to sue the other when there was no contractual nexus between them? Was Lumley's action motivated by spite or ill will? Was this why some members of the court introduced the idea of malicious behaviour as an element of liability in this case? In effect, was the court making a rule to govern the precise situation in *Lumley v Gye*, couching it in language that was of more general application? Was this, therefore, an example of hard cases making bad law?

Reading between the lines, and despite the talk in the judgments about an analogy between the situation in this case and the master-servant relationship leading to legal consequences for enticement, it would seem that the court was enunciating a policy to be adopted and enforced by the law: to protect contracts like that between Lumley and Wagner and ensure that they are upheld and performed. As we now know, in later decisions

[29] *Lumley v Gye*, above n 1, at 232.
[30] *Ibid*, at 230.

these aims were extended to cover not only contracts of service or for services but all contracts, of whatever nature. As we further now know, the protection provided by the law extends (a) to cover acts that prevent the performance of a contract even though the defendant has not induced or procured one party to commit a breach, as long as he or she has brought about a breach[31] and (b) to include within its scope acts which indirectly bring about the inducing or procurement of a breach.[32] Such extensions of the original judgment are questionable. The first may be unnecessary, since the conduct involved might be actionable on other tortious grounds, such as trespass. The second seems to come close to creating liability for conduct with only a remote connection to the ultimate breach of contract. Indeed the application of this extension of *Lumley v Gye* has caused problems.

It may be surmised that Crompton, Erle and Wightman JJ little knew what a can of worms they had opened when they held that Lumley had a good cause of action against Gye. They could not foresee the confusion, recently demonstrated by Lords Hoffmann and Nicholls,[33] between the *Lumley v Gye* tort and the condemnation of interference not directly concerned with causing the breach or non-performance of a contract but rather aimed at disrupting the economic life of the plaintiff. Where the *Lumley v Gye* tort, in its original form, did not entail the use of unlawful means or acts, the newer tort made such means or acts an integral part of liability. The initial irrelevance to liability under *Lumley v Gye* of an unlawful act to bring about the breach of contract is perhaps the most remarkable aspect of the case. It seems to indicate, even underline, the tremendous legal significance of contracts. To cause a breach of contract, despite the absence of any otherwise criminal or tortious act, was sufficient for liability. Contracts were to be taken seriously.[34]

This is not to say that prior to 1853 contracts were regarded lightly—quite the contrary. The nexus between two or more parties created by a contract gave rise to an important personal relationship. Indeed the personal nature of contract was a shibboleth of the law. The doctrine of privity came to maturity in the early nineteenth century. According to this doctrine, contracts concerned the parties thereto and no one else. An

[31] *GWK Ltd v Dunlop Rubber Co Ltd* (1926) 42 TLR 376 (KB).
[32] *DC Thomson & Co Ltd v Deakin* [1952] Ch 646 (CA).
[33] In *OBG Ltd v Allan*, above n 3.
[34] Another important aspect of *Lumley v Gye* might be the emergence of the notion that the intentional causing of economic, as contrasted with physical, harm or damage could per se be the basis for liability in the absence of a legitimate reason or excuse. This is intimately connected with the idea that certain types or instances of commercial competition were not considered valid behaviour and so were not to be countenanced. A similar attitude evolved with respect to economic duress making a contract voidable, if not void. In this context, too, the courts were faced with the need to differentiate valid from invalid behaviour, namely pressure that affected the making of an agreement.

important exception, operating long before *Lumley v Gye*, revolved around the relationship of master and servant. One who prevented a servant from fulfilling his obligations to his master by wrongfully inflicting an injury on the servant could be made liable to the master for his resultant loss.[35] This did not apply except where it was a servant who was injured. It did not extend to contracts for the provision of services as distinct from contracts of service.

The reason for such liability, according to Kitto J of the High Court of Australia in *Attorney-General (NSW) v Perpetual Trustee Co*,[36] was that such conduct constituted a wrongful invasion of the quasi-proprietary right a master was considered to possess in respect of the services the servant was under an obligation to render.[37] Several arguments supporting the view of Kitto J were subsequently adumbrated by the High Court of Australia in *Zhu v Treasurer of New South Wales*,[38] following the judgment of Jordan CJ of New South Wales in *Independent Oil Industries Ltd v Shell Co of Australia Ltd*,[39] which had analysed the nature of the defence of 'justification' in relation to inducing a breach of contract in terms of a superior 'property' right.[40] That approach was approved and endorsed in *Zhu*. It is an approach that severely limits the scope of the defence of justification, confining it to inducing a breach of, or interfering with, a contract when the alleged wrongdoer was merely protecting his or her own superior right of property, as explained by the English Court of Appeal in *Edwin Hill & Partners v First National Finance Corp plc*.[41]

II. THE DEFENCE OF JUSTIFICATION

In reaching its conclusion in *Zhu* the High Court of Australia rejected the suggestion of a wider scope to justification involving protection of contractual rights equal or superior to the rights of the plaintiff. The court, like Jordan CJ earlier, following Lord Halsbury LC in *South Wales Miners' Federation v Glamorgan Coal Co Ltd*,[42] also seems to have rejected the idea that justification could be founded on moral or religious grounds or on any duty or moral right to tender advice.

In these cases the courts seem to be declaring incorrect the decision of Russell J in *Brimelow v Casson*, in which the defendants were not liable

[35] G Jones, 'Per Quod Servitium Amisit' (1958) 74 *LQR* 39. See also the discussion accompanying n 8.
[36] (1952) 85 CLR 237 (HCA) 294–5.
[37] *Cf Lumley v Gye*, above n 1, at 232 (Erle J).
[38] (2004) 218 CLR 530 (HCA) 573–7 [*Zhu*].
[39] (1937) 37 SR (NSW) 394.
[40] *Cf* Bagshaw, above n 2, at 735.
[41] [1989] 1 WLR 225 (CA).
[42] [1905] AC 239 (HL) 244–5.

because they had acted out of a sense of morality in order to prevent actresses employed by the plaintiff from having to resort to prostitution to make up for their meagre salaries.[43] If Lord Halsbury, Jordan CJ and the High Court of Australia are correct, it follows that, as far as liability under *Lumley v Gye* is concerned, morality is irrelevant. It, in Koko's words, 'has nothing to do with the case'.[44] Only legal, proprietary or perhaps quasi-proprietary rights are pertinent.

If this is indeed the law I find it appalling. Even John Austin, I venture to suggest, would be upset. In this regard I should refer to a recent commentator who wrote that: 'The case against *Lumley* is that it assumes that there is something inherently wrong with persuading someone to breach a contract, whereas such persuasion can often be morally justified, as well as being economically desirable.'[45] This commentator would like the law to take a broader, more liberal approach to justification, so as to confine the operation of *Lumley v Gye*. He advocates a 'maximalist approach' which would require consideration of whether it was 'fair, just and reasonable' for *Lumley v Gye* to apply, which would tend to make the operation of the tort somewhat akin to the operation of the tort of negligence. Were this approach to be accepted, the courts would have to determine whether (a) in order to succeed in a claim, the plaintiff would have to show that it was fair, just and reasonable to allow the action, or (b) to avoid liability, the defendant would have to prove that to allow the action would not be fair, just and reasonable. In other words, is justification a defence or is its absence an essential ingredient of prima facie liability?

Regardless of the ultimate resolution of this question, the meaning of justification as a response to an allegation of inducing a breach of contract, in light of the explanation referred to above and the moral approach invoked in *Brimelow v Casson*, requires that reference be made to the language of Romer LJ in *Glamorgan Coal Co Ltd v South Wales Miners' Federation*.[46] He said that whether interference was justified depended on:

> the nature of the contract broken; the position of the parties to the contract; the grounds for the breach; the means employed to procure the breach; the relation of the person procuring the breach to the person who breaks the contract; and . . . the object of the person in procuring the breach.[47]

This language, which might be interpreted as opening the way to justifying inducement of a breach of contract on moral grounds, was cited and

[43] [1924] 1 Ch 302.
[44] WS Gilbert and A Sullivan, 'The Mikado' in *The Annotated Gilbert and Sullivan—Vol 1* (London, Penguin, 1982) Act II, line 646.
[45] Howarth, above n 24, at 224.
[46] [1903] 2 KB 545 (CA), aff'd [1905] AC 239 (HL).
[47] *Ibid*, at 574.

approved, but interpreted differently, by Gale CJ in *Posluns v Toronto Stock Exchange*.[48] In his view there had to be a legal basis for intervention, such as a statute or contract, or some significant social reason. A moral basis for intervention would not be enough unless joined with the promotion of a legitimate trade union or other interest that advanced a social policy. The language of Gale CJ, I suggest, might save *Brimelow v Casson* from being overruled. But the language of Jordan CJ, or the judgment in *Zhu*, and the language of English judges Darling J and Buckley LJ, cited and discussed in *Zhu*, will not.

The search for the meaning of justification in regard to the *Lumley v Gye* tort, the effect of which would be to limit materially the scope of such liability and to control the extent to which the law protects otherwise valid contracts, may arguably be said to still be ongoing. In this respect, I think, enough has been said to indicate that the modern development of *Lumley v Gye* raises serious questions as to the role of morals and morality in the determination of legal doctrine.

This is not the place to discuss that issue. My purpose in pursuing this examination of the background to, and effects of, the decision in *Lumley v Gye* was to consider the case from the point of view of judicial activism. I can only hope that I have managed to reveal that what was done in 1853 was something that was contrary to precedent and a step in the wrong direction, and, in the end, was capable of producing undesirable consequences. The point was made, at the time, by Coleridge J, who should have the last word:

> It seems to me wiser to ascertain the powers of the instrument with which you work, and employ it only on subjects to which they are equal and suited; and that, if you go beyond this, you strain and weaken it, and attain but imperfect and unsatisfactory, often only unjust, results. But, whether this be so or not, we are limited by the principles and analogies which we find laid down for us, and are to declare, not to make, the rule of law.[49]

III. ADDENDUM

Although not strictly necessary to the preceding argument, some information about the personalities involved as counsel or judges in *Lumley v Gye* is not without interest.

About counsel for the plaintiff, Cowling, I know nothing. He does not appear to have been one of the leaders of the Bar, nor did he achieve appointment to one of the common law courts. Very different was the career of counsel for the defendant, James Willes. He was described by Sir

[48] (1964) 46 DLR (2d) 210 (Ont HC) 270–73, aff'd (1965) 53 DLR (2d) 193 (Ont CA), aff'd [1968] 1 SCR 330, (1968) 67 DLR (2d) 165.

[49] *Lumley v Gye*, above n 1, at 269.

William Holdsworth as 'perhaps the most learned common lawyer of his day'.[50] Not long after he appeared for the defendant in *Lumley v Gye* he was appointed a Justice of the Court of Common Pleas. As a judge he was responsible for some highly important decisions, in which he revealed a propensity for innovation that is to be contrasted with his affirmation of the status quo in his argument in *Lumley v Gye*.[51] Undoubtedly he would have attained higher judicial office. Unfortunately he suffered from heart disease, gout and insomnia, and in 1872, following a nervous breakdown after a heavy assize at Liverpool, he shot himself.

Justice Crompton, after practising in the Court of Exchequer, became a judge of the Court of Queen's Bench in 1852. There he proved himself to be, according to Holdsworth, 'a very sound lawyer and a good judge'.[52] Justice Erle moved after two years in the Court of Common Pleas to the Court of Queen's Bench in 1847, but returned to the former court in 1859 as Chief Justice. Several years later he resigned and became a member of the Trade Unions Commission. His influential book on the law relating to trade unions was based on his memorandum to the report of that Commission.[53] Justice Wightman, notwithstanding his exalted position as a judge of the Queen's Bench, never lost his innate modesty. He had a profound knowledge of the law and possessed the essential judicial qualities of patience in listening, discrimination in judging, and clearness in explaining.[54] Justice Coleridge, grand-nephew of the poet, had a distinguished career, first as an academic and then at the bar, culminating in his appointment as a Justice of the Court of King's Bench. Currently there is another Coleridge, presumably a descendant, sitting as a judge of the High Court in England. What he thinks of his ancestor's dissent in *Lumley v Gye* I do not know. It might be interesting to find out.

[50] Holdsworth, above n 16, at 506. See also *ibid*, at 505–508.

[51] See, eg, *Collen v Wright* (1857) 8 E & B 647, 120 ER 241; *Indemaur v Dames* (1866) LR 1 CP 274; *Barwick v English Joint Stock Bank* (1867) LR, 2 Ex 259; *Austin v Dowling* (1870) LR 5 CP 534.

[52] W Holdsworth, *History of English Law* (London, Methuen & Co Ltd, 1952) vol 13, 437. His judgment in *Lumley v Gye*, however, makes this assertion questionable.

[53] Holdsworth, above n 16, at 454.

[54] E Foss, *The Judges of England* (New York, AMS Press Inc, 1966) vol 9, 203.

10

Contracting Out of Liability for Deceit, Inadvertent Misrepresentation and Negligent Misstatement

MARK P GERGEN

THIS ARTICLE EXAMINES when an actor may contract out of liability for misleading another. I take a broad perspective, covering cases in which an actor negligently or innocently misleads another as well as cases in which an actor knowingly misleads another. My principal focus is on the ability of an actor to disclaim liability for a misrepresentation made to induce a contract. I also will look at the ability of an actor to disclaim liability for misleading a claimant in the claimant's dealings with a third person. The topic straddles the law of contract, tort, unjust enrichment and equity. In tort, it encompasses the tort of fraud or deceit and the tort of negligent misstatement.[1] There are two limitations on scope. I exclude cases in which claimants suffer bodily harm or physical harm to their own property as a result of being misled. And I exclude cases in which an actor harms a claimant by misleading a third person. I focus on United States law.

While the law is unsettled on a few important issues, some basic points are fairly well established. An actor generally is strictly responsible for the accuracy of a representation made to induce a contract. To avoid responsibility, an actor must warn the other expressly not to rely or, if not that, then disclaim liability expressly should a representation turn out to be inaccurate. An exculpatory agreement will absolve an actor from liability for an inadvertent misrepresentation, including both an innocent and a

[1] Americans call it the tort of negligent misrepresentation. I will follow the practice in the rest of the common law world of calling the tort negligent misstatement, reserving the term misrepresentation for a misstatement regarding a contract between the actor and the claimant.

negligent misrepresentation (though there is some confusion on negligence).[2] There is significant disagreement over when, if ever, and why an exculpatory agreement will preclude a fraud claim. Part I examines this issue in some detail. It assesses the reasons for enforcing exculpatory agreements and concludes that the arguably valid reasons, which generally involve protecting people from baseless accusations of fraud, should be adequately served by a rule that requires fraud to be proved by clear and convincing evidence.

Turning from misrepresentation to misstatement, the key point is that generally an actor is not legally responsible for the accuracy of information which her or she disseminates when the information is not supplied as an inducement to contract or other transaction benefiting the actor,[3] though an actor can never knowingly mislead others.[4] Generally, an actor who disseminates information has a duty of care to a recipient only if her or she appears to intend to invite the recipient to rely on the information in making a decision. The requirement for a duty of care of invited reliance gives people the ability to define when they have a duty of care in supplying information, and the content and scope of this duty. On the other hand, if an actor does invite a recipient to rely on information, then an exculpatory agreement will absolve the actor from negligence liability only if the exculpatory term is clear and reasonable in the circumstances.[5]

[2] See below Part II.

[3] I do not examine the ability of people to absolve themselves from liability for misleading a claimant in the claimant's non-contractual dealings with them. Misrepresentation is a legal basis for reversing a gift but rarely will donees attempt to absolve themselves from responsibility for misleading a donor. The problem may arise in the law of equitable estoppel, which will prevent people from asserting a right, claim or defence if it would be inequitable for them to do so given that they misled the claimant to believe they would not assert the right, claim or defence. I explain the relationship of this body of law to cognate doctrines in the law of contract, tort, and unjust enrichment in MP Gergen, 'Towards Understanding Equitable Estoppel' in C Rickett and R Grantham (eds), *Structure and Justification in Private Law: Essays for Peter Birks* (Portland, Hart Publishing, 2008).

[4] Liability for deceit is not limited to misrepresentation inducing a contract with the actor. There is little authority on the power of people to absolve themselves from liability for deceiving a claimant when they supply information to a claimant regarding a transaction with a third person. The issue will arise only when an actor supplies information pursuant to a contract with a claimant or in a form that can contain exculpatory terms.

[5] See below Part III.

I. CONTRACTING OUT OF LIABILITY FOR DECEIT

There is an old saying that 'fraud vitiates all contracts'.[6] This is not strictly true. The right not to be deceived must be alienable or waiveable.[7] People routinely submit to the risk of being deceived when they play games that involve deception. There may be a few situations in which deception is an accepted part of commerce. A liar's loan may be a recent example.[8] And there are less exotic situations in which it is in a person's interest to waive the right to legal redress for fraud in the inducement of a contract. In theory, this is possible whenever the expected value of having this right is less than the expected cost to the other party to a contract. If it may be in a person's interest to waive the right to legal redress for fraud, and a person knowingly agrees to a waiver, then it is difficult to make a case for not enforcing the waiver. I will return to this point below when I examine consent as a possible basis for enforcing an exculpatory agreement.

A variation on the old saying gets us close to its true meaning. The variation is that 'no form of contract can stand, if induced by fraud'.[9] The reason is that

> the ingenuity of draftsmen is sure to keep pace with the demands of wrongdoers, and if a deliberate fraud may be shielded by a clause in a contract that the writing contains every representation made by way of inducement, or that utterances shown to be untrue were not an inducement to the agreement, sellers of bogus securities may defraud the public with impunity, through the simple

[6] *Wu v Chang* 823 A 2d 1197 (Conn 2003); *Bogosian v Bederman* 823 A 2d 1117 (RI 2003); *Snyder v Lovercheck* 992 P 2d 1079 (Wyo 1999); *Albany Urology Clinic, PC v Cleveland* 528 SE 2d 777 (Ga 2000) (in a dissenting opinion).

[7] *Whalen v Connelly* 545 NW 2d 284, 294 (Iowa 1996) is an uncontroversial example. The claimant learned of the possibility that he had been defrauded in entering into a partnership agreement but agreed to enter into a renewed agreement in which he explicitly disaffirmed any fraud claim. A more commonplace and controversial example involves the effectiveness of an incontestability clause in an insurance contract. An incontestability clause prevents an insurer from denying coverage based on a misrepresentation in procuring the insurance after the insurance has been in effect for a sufficient period. Many states have laws requiring insurance contracts include incontestability clauses. Courts are split on whether or not the clause precludes a fraud claim. *Paul Revere Life Insurance Co v Haas* 644 A 2d 1098, 1108 (NJ 1994) holds that the clause does preclude denial of coverage based on fraud and states this is the majority rule. For the contrary view see *Mutual Life Insurance Co of New York v Insurance Commissioner for State of Maryland* 723 A 2d 891, 896 (Md 1999); *Estate of Doe v Paul Revere Insurance Group* 948 P 2d 1103, 1114 (Hawaii 1997); *New England Mutual Life Insurance Co v Doe* 688 NYS 2d 459, 462 (NY 1999).

[8] These are also called 'stated income' or 'no doc' loans. A loan applicant is told that one must state an amount or source of income or an amount of assets to qualify for a loan but that little or no documentation need be supplied. It was also understood the lender would not try to verify the statement.

[9] *Arnold v National Aniline & Chemical Co* 20 F 2d 364, 369 (2nd Cir 1927) (New York law).

> expedient of placing such a clause in the prospectus which they put out, or in the contracts which their dupes are asked to sign.[10]

There is a narrow proposition here that is uncontroversial. It is that an exculpatory term—for example, a disclaimer, a merger provision, or a non-reliance clause—will not shield fraud if it is in a document which a claimant is unlikely to have read or to have understood. Courts that enforce exculpatory agreements exclude terms in form contracts or boilerplate.[11] A related qualification that is universally recognised is that an exculpatory term cannot shield a contract from a challenge on the ground that consent to the term was obtained by fraud or trickery.[12] Some courts that enforce exculpatory agreements impose further limitations. Some require that the claimant had read the contract and had been represented by a lawyer.[13] Some hold a merger provision will preclude a fraud claim on an extra-contractual promise but not a fraud claim on an extra-contractual misrepresentation of fact.[14] Some hold a disclaimer will not shield an actor from a fraud claim on a representation involving 'facts . . . peculiarly within the knowledge' of the actor.[15]

Danann Realty Corp v Harris[16] illustrates the sort of case in which the issue of the enforceability of an exculpatory agreement is up for grabs.[17] It is the leading New York case on the subject. The claimant purchased a long-term lease allegedly relying on oral representations by the seller regarding the property's operating expenses and profits. The contract included terms stating the seller had made no representations regarding expenses, the buyer had made a personal inspection, the contract was the

[10] *Ibid*.

[11] *Danann Realty Corp v Harris*, 157 NE 2d 597 (1959) [*Danann Realty*] limited the rule to exclude a 'general and vague merger clause' to distinguish *Sabo v Delman* 143 NE 2d 906 (1957), a case from two years earlier that held an exculpatory term would not preclude a fraud claim. WE Nelson, 'From Morality to Equality: Judicial Regulation of Business Ethics in New York, 1920–1980' (1999) 43 *New York Law School Law Review* 223, 276–7 attributes the embrace of exculpatory agreements in *Danann Realty* to a shift on the court to a majority that favoured business efficiency over business ethics. A later New York case holds a negotiated merger clause will also preclude a fraud claim on an extra-contractual representation: *Citibank v Plapinger* 485 NE 2d 974 (NY 1985).

[12] *Toy v Metropolitan Life Insurance Co* 928 A 2d 186 (Pa 2007); *Yocca v Pittsburgh Steelers Sports Inc* 854 A 2d 425 (Pa 2004); *Hamade v Sunoco Inc (R & M)* 721 NW 2d 233 (Mich App 2006).

[13] See *LaFazia v Howe* 575 A 2d 182 (RI 1990) (giving effect to an exculpatory term when the claimant read the contract and was represented by a lawyer).

[14] *Watkins & Son Pet Supplies v Iams Co* 254 F 3d 607 (6th Cir 2001) (Ohio law).

[15] *Aetna Cas & Surety Co v Aniero Concrete Co Inc* 404 F 3d 566, 575–76 (2nd Cir 2005); *Banque Arabe et Internationale D'Investissement v Maryland National Bank* 57 F 3d 146, 155 (2nd Cir 1995).

[16] Above n 11.

[17] For the position that exculpatory terms cannot preclude a fraud claim, see, eg, *Lusk Corp v Burgess* 332 P 2d 493 (Ariz 1958); *Aspiazu v Mortimer* 82 P 3d 830 (Idaho 2003); *First National Bank in Durant v Honey Creek Entertainment Corp* 54 P 3d 100 (Okla 2002); *Shah v Racetrac Petroleum Co* 338 F 3d 557 (6th Cir 2003) (Tennessee law).

complete agreement, and neither party was relying on representations outside the contract. The New York Court of Appeals held the contract precluded a fraud claim by treating the exculpatory terms, particularly the non-reliance clause, as conclusive proof of the absence of fraud. I will come back to the arguments for enforcing an exculpatory agreement after I examine how courts that refuse to enforce exculpatory agreements handle fraud claims.

A. 'No Form of Contract Can Stand, If Induced by Fraud'

The proposition 'no form of contract can stand, if induced by fraud', if taken literally, would mean it was impossible to write a contract that will preclude a party from challenging the contract by claiming fraud in the inducement. This might seem to belie my claim earlier that the right not to be deceived is alienable or waiveable. If it truly were impossible to write a contract to preclude a claim of fraud in the inducement, then the right not to be deceived regarding a contract in effect would be inalienable and unwaiveable.

While this may be the practical effect of the rule, none of the reasons for the rule requires denying the possibility that a person could agree to bear the risk of being deceived. The rule is grounded largely on the observation that people often 'accept . . . and act upon agreements containing . . . exculpatory clauses . . . but where they do so, [they are] nevertheless, in reliance upon the honesty' of the actor.[18] The explanations for this sort of behaviour are familiar. When an exculpatory term is in boilerplate in a form contract it probably is unread. Because of the strong expectation of honesty, even if people do read an exculpatory term, they may not think it covers dishonesty, or they may underestimate the risk of dishonesty.[19] But these are merely reasons to be sceptical about the quality of assent to exculpatory agreements. Reasons of this type cannot establish that it could never be in a person's interest to agree to bear the risk of being deceived. The other argument for the rule is that the law cannot spell out when exculpatory terms will and will not be enforced for the cunning will exploit

[18] *Snyder v Lovercheck,* above n 6, at 1086. And, we might add, they would not think they had agreed to relinquish the right to seek legal redress for deceit.

[19] JM Lipshaw, 'Of Fine Lines, Blunt Instruments, and Half-Truths: Business Acquisition Agreements and the Right to Lie' (2007) 32 *Delaware Journal of Corporate Law* 431, makes an argument along these lines to justify a rule construing exculpatory terms narrowly. R Prentice, 'Contract-Based Defenses in Securities Fraud Litigation: A Behavioral Analysis' [2003] *University of Illinois Law Review* 337, argues that often when people do agree to exculpatory terms it is because they irrationally underestimate the likelihood that the person with whom they are dealing is acting dishonestly.

any such rule (or at least any feasible rule[20]) to defraud the unwitting. But this does not foreclose the possibility of enforcing an exculpatory agreement. It only forecloses having rules to define when an exculpatory agreement will be enforceable.

But neither are there strong reasons to reject the rule. The best reason for enforcing exculpatory terms is to protect innocent people from baseless accusations of fraud. A persistent worry in the United States—attributable in significant part to our reliance on the jury to resolve contestable issues of fact (particularly when a fact issue turns on credibility), to distrust of the jury, and to the high cost of litigation—is that people who are disappointed with a contract, but who have no contractual grounds for complaint, will claim fraud on slight evidence in the hope of finding a sympathetic jury or in the hope that the high cost of litigation will prompt the other party to settle on favourable terms. Rules requiring that fraud be pled with specificity and be proven by clear and convincing evidence are safeguards against this sort of behaviour. They enable courts to police fraud claims and direct courts to resolve factual doubts against a claimant. What an exculpatory agreement will not do is to shield an actor from having to defend himself against a fraud claim if the claimant can make out a plausible case. And an exculpatory agreement will not shield an actor from liability for fraud if there is strong evidence.

Snyder v Lovercheck shows how the approach works in practice.[21] Snyder purchased the Loverchecks' wheat farm during the winter after inspecting the farm numerous times. Many of the visits were in the company of Snyder's agent (Hayek) and the seller's agent (Ron Lovercheck). Snyder alleged that on the first visit he was told by Ron that 100 of the 1,960 acres had a problem with rye, which reduces wheat production. Later Ron told Snyder that he had confirmed this by talking to a prior owner. The sales agreement, which was a form contract published by the state real estate commission, said the farm was sold 'as is' and that the buyer was not relying on representations made by the seller or the seller's agents. When the first crop came in Snyder discovered that the rye's presence was much worse than he had been told and that it reduced the value of the land by one-quarter of the purchase price. He brought suit asserting breach of warranty, negligent misstatement, and fraud. He also sued his agent, Hayek, asserting negligence for failing to warn him about

[20] I can imagine a rule giving effect to an exculpatory term that would be difficult for the cunning to exploit. For example, the law might require an earnest statement from the claimant to the effect: 'I know you may lie to me and I agree to bear that risk.' But it is difficult to imagine someone asking for such a concession in the usual settings in which it is in a person's interest to agree to waive the right to legal redress for fraud. One such setting is where a person has good reason to trust an actor but recognises others who do business with the actor may not be as trusting and may falsely accuse the actor of fraud.

[21] Above n 6.

the exculpatory terms. The trial court granted summary judgment for all defendants on all claims and awarded costs and legal fees to the sellers and their agent. I will return to the court's handling of some of the other claims later.

On appeal the Loverchecks and Ron sought to preserve their victory by urging the court to adopt the New York rule enforcing exculpatory agreements. The court declined to do this. But the court's rejection of the New York rule did not lead it to remand the case for a trial on the merits of the fraud claim. The court concluded summary judgment was appropriate, stating as a rule 'fraud will not be imputed to any party when the facts and circumstances out of which it is alleged to arise are consistent with honesty and purity of intention'.[22] In other words, dishonesty will not be imputed if there is a reasonable chance an actor did not know a representation was false (and was not reckless with regards to its accuracy), or did not intend or have reason to expect that the claimant would rely on the representation.[23]

I think the Wyoming court got it right both in rejecting the New York rule and in adopting rules that require resolving doubts about the existence of fraud against a claimant. The policy strikes a decent balance between the competing goals, which generally are deterring and redressing deceit while protecting honest actors from baseless accusations of deceit. This also achieves the purpose of most exculpatory agreements. And by rejecting the New York rule the court avoided the difficulties of defining when and explaining why an exculpatory agreement will shield fraud. I turn to these difficulties now.

B. Enforcing Exculpatory Agreements

Often when courts enforce an exculpatory agreement it is to dismiss a weak claim that could also have been dismissed under rules that require resolving doubt about the existence of fraud against a claimant. But there are cases in which an exculpatory agreement seems to have decisive effect.

[22] *Ibid*, at 1086.

[23] *Cushman v Kirby* 536 A 2d 550 (1987) illustrates the sort of facts that suffice to establish deceit in a similar situation. This was a sale of a home 'as is' in which the wife of the couple selling the home responded to an inquiry about the well water 'It's good. It's fine.' Her husband stood silently by while she said this. The buyers did not taste the water. When they moved in they discovered the water had a strong sulphur taste. The buyers were allowed to recover the cost of hooking up to the municipal water supply on grounds of fraud. The husband admitted he was aware of the problem and that he heard his wife's statement. The jury was allowed to infer that he remained silent hoping to induce the buyers to purchase the home. As for the wife, the jury was allowed to infer that she knew her statement was partial and misleading and she made it hoping to induce the buyers to purchase the home. *Snyder v Lovercheck* would be similar if the sellers had known of their broker's representation and they had known from prior crops that the rye's presence was much worse than represented.

LaFazia v Howe is an example.[24] The Howes, an elderly couple, purchased a delicatessen from LaFazia and Gasrow. The Howes' court papers tell a plausible story of deceit. They alleged the sellers told them the deli grossed $450,000 to $500,000 per year. The Howes initially were dubious because the seller's tax returns showed a much lower income. The sellers explained this away by pointing to their fancy houses and fancy cars and their lack of another source of income. Soon after they took over the deli the Howes discovered its income was a fraction of what had been represented. They tried to make a go of it but after six months they sold the deli for half of what they had paid for it. The contract contained strong terms disclaiming representations and reliance and stating the deli was sold 'as is'. The trial judge granted the sellers' motion for summary judgment. He admonished the Howes from the bench that they had seen 'the tax returns didn't justify the asking price' and that the contract had warned them they were 'making their own judgment' and 'acting upon their own'. The Rhode Island Supreme Court affirmed, holding that a clear exculpatory term precludes a fraud claim. In the rest of this part I examine the arguments for enforcing exculpatory agreements. I conclude they do not justify the result in *LaFazia v Howe*, and generally do not justify enforcing an exculpatory agreement when there is clear and convincing evidence of fraud, with the possible exception being an agreement absolving an actor from liability for third party fraud.

(i) Some Clearly Bad Arguments

Courts like those in New York and Rhode Island that enforce exculpatory agreements struggle to square this result with the law of deceit. This is difficult to do because the law of deceit is designed to protect people like the Howes, who are naïve, stupid or too trusting for their own good, from people like LaFazia and Gasrow. The so-called 'Double-Liar' argument is the worst of the bunch.[25] The gist of the argument is that claimants can assert a fraud claim only if they lied in representing that they did not rely on a representation. Or as one judge has said: 'To allow the buyer to prevail on its claim is to sanction its own fraudulent contract. The enforcement of non-reliance clauses recognises that parties with free will should say no rather than lie in a contract.'[26]

[24] Above n 13.

[25] The argument appears in *Danann Realty*, above n 11. People who should know better have made the argument. It is made by Judge (now Justice) Alito in *MBIA Insurance Corp v Royal Indemnity Co* 426 F 3d 204, 218 (3rd Cir 2005), by Judge Easterbrook in *Rissman v Rissman* 213 F 3d 381 (7th Cir 2000), and by Vice Chancellor Strine in *ABRY Partners v F&W Acquisition, LLC* 891 A 2d 1032 (Del Ch 2006). [*ABRY Partners*].

[26] *ABRY Partners*, *ibid*, at 1058.

The silliness of the argument is clear if we consider its implications in *LaFazia v Howe*. The argument rests on the dubious factual premise that the non-reliance clause was understood by the Howes to be a meaningful representation of their state of mind in buying the deli. Even conceding this premise, the representation would not be fraudulent if the sellers knew it was false—that is, if the sellers knew the Howes were relying on their representations notwithstanding the non-reliance clause. If the Howes did anything that might be dishonest, then it is by asserting a fraud claim to get around defences to a contract claim. But it is dishonest for the Howes to claim fraud only if they knew the contract was meant to absolve the sellers from liability for deceit. And even then this would be dishonest only if the Howes were insincere in agreeing to absolve the sellers from liability for deceit. We would have to imagine that the Howes anticipated that the sellers might be lying, that the Howes understood the contract purported to absolve the sellers from liability in the event they were lying, and that the Howes slyly agreed to the contract secretly intending to sue the sellers for fraud if it turned out the sellers were lying. People do not behave in this way.

In rejecting the Howes claim, the trial judge admonished them that they had seen the tax returns and that the contract had warned them 'they were acting on their own'. Perhaps the trial judge dismissed the claim because he thought the Howes had acted unreasonably. A few cases hold that a clear exculpatory term renders a claimant's reliance unreasonable, absolving the defendant from liability for having misled the claimant.[27] But this flies in the face of the long-standing principle that contributory negligence is not a defence to fraud.[28] I will come back to the related question of when an exculpatory agreement may render a claimant's reliance 'unjustifiable', which is an element of fraud under United States law.

A recent decision by the Delaware Court of Chancery in *ABRY Partners v F&W Acquisition, LLC*[29] stakes out a middle ground between the position that a contract can never shield fraud and the position that a clear exculpatory agreement should be enforced so long as it is not procured by fraud or trickery. Unfortunately, while the court reached a defensible result, the opinion, which is by Vice Chancellor Strine, makes a muddle of the problem conceptually. The claimant ABRY bought the stock of a publishing company from its owner, a private equity firm, for $500 million. The acquisition agreement had a term that is standard in such agreements stating that all representations were made by the target company and not by the seller. The agreement also limited the seller's obligation to a $20 million indemnity expressly stating 'misrepresentation'

[27] *Masingill v EMC Corp* 870 NE 2d 81 (Mass 2007), which I discuss later, is an example.
[28] See *Restatement (Second) of Torts* (1977) § 545A, Illustrations 1–3.
[29] Above n 25.

to be among the covered grounds. After the closing ABRY discovered facts that led it to believe that the target's financial information had been manipulated to inflate the target's value by $100 million and that the seller was complicit in the manipulation.

The case holds that a contract cannot absolve a seller from liability for fraud for representations found within the contract if the seller knew the representations were false. This leaves three exceptions covering situations in which exculpatory terms may be enforced. All are limited to 'negotiated commercial contracts' involving 'sophisticated parties'. First, a clause disclaiming reliance will absolve the seller from extra-contractual representations (that is, representations not repeated within the contract). This is similar to the rule in *Danaan Realty*. Secondly, a contract may absolve an actor from liability for a reckless misrepresentation.[30] Thirdly, a contract may absolve an actor from liability for dishonesty of another person so long as the actor is unaware of the dishonesty. In the situation of ABRY this meant the seller could absolve itself from liability for fraud by the target company.[31]

Much of this flows from two opposing principles set up by Vice Chancellor Strine in the opinion. One principle abhors dishonesty. This principle is so strong, according to Vice Chancellor Strine, that an agreement absolving an actor from liability for dishonesty is void as against public policy.[32] The other principle favours private ordering. The general rule and two of the exceptions follow straightforwardly from the two principles. The rules enable people to agree to absolve an actor from liability for fraud except in situations involving personal dishonesty. The principle favouring private ordering holds until it runs square into the principle abhorring dishonesty.

I do not think answers can be found reasoning from these particular principles. The two principles do not explain why an exculpatory agreement should absolve an actor from liability for a representation not repeated within a contract (policy reasons are the best explanation for being extra-sceptical about alleged oral representations). While there is some merit in the distinction between actors absolving themselves from liability for personal dishonesty and actors absolving themselves from

[30] Fraud usually encompasses a reckless misrepresentation, meaning a representation made with reckless indifference to its possible falsity. A related concept defines as being fraudulent averring a fact as true with knowledge there is not adequate basis for ascertaining the truth of the fact.

[31] The third rule could develop into a significant exception if it is extended to permit a principal to absolve himself from liability for an agent's fraud.

[32] *ABRY Partners*, above no 25, at 1062. Vice Chancellor Strine even suggests such a rule may be justified on efficiency grounds. If the detection and correction of fraud was costless and error-free this generally would be true (putting to the side games involving deception). Of course, it is not, and so it is not clear that it could never be in the joint interests of contracting parties to preclude a fraud claim.

liability for dishonesty by a third party,[33] Vice Chancellor Strine draws too sharp a line. The primary reason to enforce exculpatory terms, which is to protect honest actors from baseless accusations of fraud, applies to claims of personal dishonesty. The primary reason not to enforce exculpatory terms, which is scepticism about the quality of assent, applies to claims of third party dishonesty. Indeed, much fraud is committed by agents. There also is some merit in the distinction between knowing and reckless misrepresentation. At a fundamental level inadvertent misrepresentation is different from intentional misrepresentation both as a matter of morality and as a matter of policy. But intentional, knowing and reckless misrepresentation are difficult to distinguish, particularly once one grapples with problems of proof.

(ii) Policy

Judge Augustus Hand aptly summarised the policy argument for enforcing exculpatory agreements in 1927, ascribing the view to Massachusetts courts (which later recanted):

> The Massachusetts cases emphasize the desirability of certainty in the contractual relations of those who have made a definite agreement, and if they say that they contract without regard to prior representations and that prior utterances have not been an inducement to their consent, any occasional damage to the individual caused by antecedent fraud is thought to be outweighed by the advantage of certainty and freedom from attacks, which would in the majority of cases be unfounded where such provisions were in the agreement.[34]

The key claim is in the final words. This is a claim that the absence of fraud can be inferred from the presence of an exculpatory agreement. This claim is wildly unrealistic if one takes the entire universe of contracts. But the rule enforcing exculpatory agreements excludes form contracts and boilerplate and is limited to what may loosely be described as negotiated commercial contracts. An exculpatory agreement has some evidentiary value in such a contract. But a rule enforcing an exculpatory agreement goes further and treats it as conclusive evidence of the non-existence of fraud, preventing a court from considering other evidence no matter how strong it might be. So in *Danann Realty* the court said 'a specific disclaimer destroys the allegations in the complaint that the agreement was executed in reliance upon these contrary representations'. A charitable interpretation of this is that a disclaimer 'destroys' reliance by conclusively establishing its non-existence.

[33] See below Part I (B)(iv).

[34] *Arnold v National Aniline & Chemical Co,* above n 9, at 20. Judge Hand also cites to 25 *Columbia Law Review* 231, which is a brief case comment.

The argument for this is not entirely evidentiary. American courts and scholars have long argued over the related question whether a merger provision in a contract should be treated as conclusive evidence that the contract was indeed a final and complete statement of the parties' agreement. The view that it should be, which is associated with Williston, is best justified not by the evidentiary value of a merger provision but instead by other purposes served by treating a merger provision as conclusive. These include encouraging people to put their agreements in writing, simplifying contract enforcement, making interpretation an issue for the judge and not the jury, and strengthening the hand of appellate judges in policing contract interpretation.[35]

Connecting the evidentiary and institutional arguments are the fears that people over-claim fraud and that courts do a poor job of distinguishing good claims from bad. If these fears are borne out by reality, then a policy of enforcing exculpatory agreements would permit some fraud to go uncorrected while reducing the costs of administering contracts across many cases, including a significant number of cases in which innocent people would incur the cost of defending against unfounded fraud claims, and perhaps even eliminating some outcomes in which innocent people would be unjustly found liable for fraud.

The argument for enforcing exculpatory agreements rests on a very pessimistic view of human litigiousness and of the fallibility of legal institutions and a very optimistic view of private ordering through contract. But even if one holds these views it is not clear the rule is justified; it is a clumsy tool to discourage unfounded fraud claims and avoid unjust fraud verdicts. The rule requires a court to dismiss a fraud claim no matter how strong the evidence. The rules requiring that fraud be pled with specificity and proven by clear and convincing evidence serve the same ends while enabling a court to override an exculpatory agreement if there is strong evidence of fraud. One has to assume a very bleak view of human litigiousness and of the fallibility of legal institutions to justify enforcing an exculpatory agreement on evidentiary and institutional grounds in the face of clear and convincing evidence of fraud. So on to the next reason.

[35] JD Calamari and JM Perillo, 'A Plea for a Uniform Parol Evidence Rule and Principles of Contract Interpretation' (1967) 42 *Indiana Law Journal* 333 nicely capture this at 341 in describing the gist of the disagreement between Williston and Corbin: 'Professor Corbin has an easy task in demolishing the Willistonian approach. In treating the matter of integration as a question of intent, as Professor Williston purports to do, he shows the absurdity of excluding all relevant evidence of intent except the writing itself. But … Williston … [is] unconcerned about the true intention of the parties . . . [Williston is] advocating and applying a rule of form.'

(iii) Unjustifiable Reliance

The actual doctrinal basis for the decision of the Rhode Island Supreme Court in *LaFazia v Howe* was that the exculpatory agreement made the Howe's reliance 'unjustifiable'. The law of unjustifiable reliance may seem an unpromising place to look for a rationale for enforcing exculpatory agreements. Dan Dobbs' treatise concludes that the concept of unjustifiable reliance has no analytical content in the law of fraud. He argues that courts use the requirement of justifiable reliance as a tool to screen out claims which are weak on some other ground (that is, a weakness in proof of the fact of a misrepresentation, or of a claimant's reliance, or of an actor's dishonest intent).[36] The requirement of justifiable reliance is worrisome for it is easy to make the mistake of equating justifiable reliance with reasonable reliance. An additional reason for scepticism about the value of the concept is that it appears only in United States law. Other legal systems seem to do fine without it.

The confusion of unjustifiable reliance with unreasonable reliance is understandable. Much of the law of unjustifiable reliance is rules defining when an actor who may have knowingly or recklessly supplied inaccurate information to a claimant is not liable for fraud because the actor probably did not intend, or have reason to expect, the claimant would rely on the information because a reasonable person in the claimant's situation would not rely on it. The mens rea of fraud requires that an actor intend, or have reason to expect, that a claimant would attach significant importance to a representation.[37] This is in addition to the requirement that an actor know a representation is inaccurate or be reckless with regards to its accuracy. The rules precluding liability for opinion and prediction are best explained on this basis. A salesman who puffs his wares may not intend or have reason to expect a buyer will attach importance to the puffery in deciding whether to buy because most people discount puffery.

Often the issue is amenable to being resolved by rule. When an actor's actual intent is unknown, liability will turn on whether the actor had reason to expect the claimant would attach importance to a representation. This in turn depends on whether a normal or reasonable person would be expected to do so, unless an actor has particular reason to know a claimant

[36] D Dobbs, *The Law of Torts* (St Paul, West Group, 2000) 1360–61. Dobbs' diagnosis of the law of unjustifiable reliance resonates with the policy argument for enforcing exculpatory terms just discussed. Bruce Feldthusen concludes the element of justifiable reliance serves no function in the law of negligent misstatement that is not as well served either by the test of pecuniary interest for duty or by the defence of contributory negligence. B Feldthusen, *Economic Negligence*, 3rd edn (Scarborough, Carswell, 1994) 64, n 166.

[37] *Restatement (Second) of Torts* (1977) § 531 is a clear statement of the general principle. The Comments to the Section focus on the issue of liability to remote and numerous claimants.

is credulous. Whether a normal person would be expected to act on a representation turns on social or general facts, and so the issue is amenable to being resolved by a rule. Again the rule on puffery is an example. The general rule treating reliance on opinion as unjustifiable is another example. In United States law most of these rules are found in the law of unjustifiable reliance. In other common law systems these rules tend to be a gloss on a rule restricting misrepresentation to statements of fact. These rules should give way if there is sufficiently strong evidence that an actor knew, or had reason to expect, a claimant was unusually credulous.

An exculpatory agreement sometimes will negate this aspect of the mens rea of fraud. *Masingill v EMC Corp* illustrates this point.[38] Masingill left a job at Compaq, relinquishing a bonus and stock options, to work for Data General. Because of moving, she worried about being protected against losing stock options in Data General if the company were bought out. Senior executives at Data General were protected against this risk by 'forward vesting'. Mid-level employees like Masingill were not. Masingill sought and received assurances from senior executives at Data General that they would work to get her 'chute' protection during the first year. But the employment contract, which was the product of much going back and forth between the parties, guaranteed her only a year's salary as severance in the event of a buyout leading to her termination. Masingill understood that this left her without the 'chute'; this was the major bone of contention in negotiating the contract.

Masingill was right to worry. EMC bought out Data General eight months into her tenure. She quit four months later when a suitable job could not be found at EMC, losing her stock options. She sued Data General and the executives individually alleging they had deceived her in assuring her that efforts would be made to get her the 'chute' during the first year. The case was tried by a jury, which found that the executives had 'knowingly and recklessly made false statements' to Massingill about getting her the 'chute'. Nevertheless a verdict was rendered for the executives. The jury was instructed by the trial court that it had to find Massingill's reliance on the representations unreasonable because the representations were explicitly contradicted by her employment contract. The Massachusetts Court of Appeals affirmed, reasoning that the contract made Massingill's reliance on the representation unreasonable as a matter of law. This basis for the decision is inconsistent with the general principle that contributory negligence is not a defence for fraud. A better rationale is that the executives probably did not intend, and did not have reason to expect, Massingill would attach importance to their assurances that they would work to get her the 'chute' given the inconsistent terms in the

[38] 870 NE 2d 81 (Mass 2007). I have simplified the facts to isolate the relevant claim.

employment contract and the negotiation history. The unreasonableness of Massingill's reliance negates the mens rea of deceit. There is an important substantive distinction between non-culpability for unexpected reliance and contributory negligence. A salesman who knows of and exploits a credulous buyer's unreasonable reliance on puffery commits fraud because the salesman knows of the buyer's credulousness.

This reasoning cannot justify a general rule enforcing exculpatory agreements even if the rule is limited to negotiated commercial contracts. A general rule elides the distinction between non-culpability for unexpected reliance and contributory negligence, if the rule requires a court to disregard evidence that an actor intended, or had reason to expect, that a claimant would unreasonably rely on a representation despite the exculpatory agreement. This does not mean that the trial judge erred in Massingill in directing a verdict for the defendants. The policy argument for enforcing exculpatory agreements justifies a directed verdict once Massingill failed to produce clear and convincing evidence that the executives were aware she continued to rely on their representations they would get her the 'chute' notwithstanding the contract terms to the contrary.

But this was not the case in *LaFazia v Howe*, and so this reasoning cannot justify the result in that case If you believe the Howes' allegation that the sellers explained away the deli's low reported earnings by implying they cheated on their taxes (and the trial judge did, for he stressed this fact in dismissing the claim), then this is clear and convincing evidence that the sellers wanted the Howes to rely on the (lying) representations of the earnings. Why else would the sellers embarrass themselves by implying they were tax cheats? The Howes' continued reliance on the representations may have been unreasonable, but that is irrelevant if the sellers intended them to rely. We still do not have a good reason for enforcing an exculpatory agreement when there is clear and convincing evidence of fraud (including evidence that the defendant intended the claimant to attach importance to the representation). So we move on to the last reason.

(iv) Consent

Consent is generally not offered as a rationale for enforcing an exculpatory agreement. *ABRY Partners v F&W Acquisition, LLC* goes so far as to reject consent as a basis for enforcing an exculpatory agreement.[39] The contract included a $20 million indemnity cap that expressly covered damages resulting from 'misrepresentation'. ABRY argued that the cap was meant to cover innocent or negligent misrepresentation but not fraud and that the cap was meant to cover claims in contract but not tort. Vice

[39] Above n 25.

Chancellor Strine properly rejected these arguments as undercut by common usage. The term 'misrepresentation' usually refers to fraud or deceit and to the tort action. But Vice Chancellor Strand refused to hold ABRY to its contract. Instead he held that an exculpatory agreement was void as against public policy except in limited situations.[40]

I can see no obvious reason in principle why consent cannot be a basis for enforcing an exculpatory agreement. As I explained earlier, the reasons for not enforcing exculpatory agreements are the suspect quality of consent and the inadvisability for having a rule defining when an exculpatory agreement is enforceable. These could be sufficient reasons to adopt a general rule of invalidity, if it never or rarely was in a person's interest to waive the right redress for fraud in the inducement of a contract. But this seems unlikely.

Vice Chancellor Strand identifies one counter-example in *ABRY Partners*. Sometimes it is very costly for a company selling a division to verify the honesty of the division's managers or the accuracy of its financial information. In this situation it makes a great deal of sense to absolve the seller from liability for fraud in the division. This puts the buyer on notice to be wary of fraud in the division and avoids the cost of duplicative anti-fraud investigations. More generally, an exculpatory agreement may be efficient[41] in a situation involving third party fraud in which an actor does not have a comparative advantage in preventing the third party fraud or in bearing the risk of the third party fraud. An exculpatory agreement may be efficient even if an actor has a comparative advantage if a claimant would take duplicative precautions in any event because the claimant cannot rely on fully recovering a loss on third party fraud from the actor.

The assumptions underpinning the policy argument for enforcing exculpatory agreements suggest a more general counter-example. The assumptions are, first, that often fraud claims are unfounded and, secondly, that courts do a poor job of distinguishing good claims from bad. The situation is that Albert has good reason to depend on Bill's honesty but A knows B deals with other people who do not trust B and who can be expected to bring an unfounded claim of fraud if they are sufficiently unhappy with the outcome of a contract. A may get a lower price by waiving his right to

[40] *Restatement (Second) of Contracts* (1981) § 196 seems to agree. It states: 'A term unreasonably exempting a party from the legal consequences of a misrepresentation is unenforceable on grounds of public policy.' The Comments do not explain what, if anything, would be a reasonable exculpatory agreement, though they do suggest a term 'reasonably' limiting the time in which misrepresentation can be asserted is enforceable. The prohibition is somewhat undercut by a statement that the rule does not apply 'to language that prevents the making of any misrepresentation in the first place, such as that disclosing the truth (see § 161). Nor does it apply to language that prevents reliance by the recipient on a misrepresentation (see § 167) or that makes his reliance unjustified (see § 172), but such language is not effective unless it actually has the asserted effect and is not a mere recital that it does.'

[41] By this I mean that it increases the joint expected return on a contract.

redress for fraud in the inducement without reducing the expected value of the contract to him.[42] More generally, an exculpatory agreement may be efficient if the expected value to the claimant of having the option to seek redress for fraud in the inducement is less than the expected cost to the actor.

The consent argument has some conceptual advantages over the other arguments for enforcing exculpatory agreements. It avoids tendentious arguments that an exculpatory agreement negates the fact of fraud. It is consistent with the practice of making the effect of an exculpatory agreement an issue for the judge with the key questions being the validity of the claimant's consent to the exculpatory terms and the apparent meaning of the terms. It resonates with the policy argument but distances us from some of its troublesome factual and normative assumptions. We can be agnostic about human litigiousness, the fallibility of legal institutions, and the relative weight to be assigned to the good and bad effects of enforcing exculpatory agreements, and defer to private choices. All we need is faith in private ordering.

Faith in private ordering is the sticking point, of course. Scepticism about the quality of consent to an agreement absolving an actor from liability for fraud remains. This scepticism requires two and maybe more limitations on a rule enforcing an exculpatory agreement. First, this is limited to negotiated commercial contracts. Secondly, there must be a plausible explanation as to why in the situation it was in the claimant's interest to waive the right to redress for fraud. If one is sceptical about the quality of consent to exculpatory agreements in general, or to the quality of consent in the particular case, then one would require more. One might enforce an exculpatory agreement only if claimants almost surely understood they were waiving the right to redress for fraud. And, other than in cases of third party fraud, one might refuse to enforce an exculpatory agreement when there is clear and convincing evidence of fraud on the reasoning that the likely purpose of an exculpatory agreement is served by shielding an actor from unfounded claims of fraud. If you come out here, then you are close to the position associated with the precept 'No form of contract can stand, if induced by fraud', but you would recognise a narrow

[42] Consider a stylised example. Sam (S) is a seller of goods that may be of low or high value. The actual value of a good is unobservable to a buyer. The actual value sometimes is known to S but S's lack of such knowledge is unverifiable. S has a well-deserved reputation for honesty. He faces buyers who differ in their willingness to trust S and in their litigiousness. S is unable to differentiate among them. S always discloses that a good is of low value if he knows it, but he knows that if he does not identify a good as being of low value he faces a risk that a buyer who does not trust him and who is litigious will successfully claim that he knew and failed to disclose a good was of low value. In this situation, buyers who trust S and/or who are not litigious may obtain a lower price by precommitting not to sue for fraudulent inducement.

exception (for third party fraud) and you would be resolute about resolving factual doubts against a claimant, particularly when there is a clear exculpatory agreement in a negotiated commercial contract.

II. CONTRACTING OUT OF LIABILITY FOR AN INNOCENT OR A NEGLIGENT MISREPRESENTATION REGARDING A CONTRACT

It is basic contract law that absent fraud actors may absolve themselves from liability for a representation regarding the subject matter of a contract by getting the claimant to agree that the representation is not part of the contract or getting the claimant to agree that claimants bear the risk of the inaccuracy of a representation. When the representation is not a term of the contract,[43] this is done by including a disclaimer, a non-reliance clause, or a merger provision in a written agreement. In the United States, the parol evidence rule is the principal legal vehicle for giving effect to such an agreement. If a court applies the strong form of the parol evidence rule—favoured by Williston and the First Restatement of Contracts[44]—then a clear exculpatory term will absolve an actor from liability for an extra-contractual representation no matter how strong the proof of the representation.[45] If a court applies the weak form of the rule—favoured by Corbin, the Second Restatement of Contracts, and the UCC—then an exculpatory term will not absolve an actor from liability for an extra-contractual representation if there is credible evidence of the representation and the court finds that the claimant reasonably believed the representation was part of the contract despite the exculpatory term.[46]

[43] When a representation is a term of a contract this is done by a limitation of remedy or some other term assigning the risk of the inaccuracy of the representation to the claimant.

[44] A good summary of the different forms of the rule may be found in LA Cunningham, 'Toward a Prudential and Credibility-Centered Parol Evidence Rule' (2000) 68 *University of Cincinnati Law Review* 269.

[45] *Mitchill v Lath* 160 NE 646 (NY 1928) illustrates this. The trial court found the seller of a house had orally promised to move an unsightly ice house he owned on adjoining property. The majority held the promise was unenforceable because it was omitted from the written contract when normally one would expect it to be included. The majority commented on the seller's 'moral delinquencies' in not fulfilling its promise and observed: '[We] have believed that the purpose behind the rule was a wise one not easily to be abandoned. Notwithstanding injustice here and there, on the whole it works for good.'

[46] *Husky Spray Service Inc v Patzer* 471 NW 2d 146 (SD 1991) illustrates this. The seller orally represented to pilots representing the buyer who test flew a used airplane that the plane was 'ready to go' and that he would repair any defects. Apparently the seller was unaware of a growing crack in a crankshaft that hampered the buyer's use of the plane and eventually ruined the engine. The sales agreement disclaimed all warranties, said the plane was being sold 'as is', and that the buyer was relying on its own inspection. Recovery was allowed on a theory of breach of warranty. The warranties were established by credible evidence (two of the pilots no longer worked for the buyer) and the seller had not specifically bargained for the disclaimers (they were on the reverse side of the contract, which the buyer testified it never read).

The differences between these rules and the rules on fraud are striking. Even the weak form of the rule assumes that an actor may disclaim liability for an extra-contractual representation regarding a contract by getting the other party to agree that the representation is not part of the contract. It also assumes that an apparent agreement will suffice. Claimants will lose even though they believe a representation is actionable if this belief is unreasonable in the circumstances. The difference between the weak form and the strong form of the rule is that under the strong form of the rule a clear exculpatory term is conclusive on the issue of apparent assent while under the weak form of the rule a court considers other facts and circumstances and decides whether, taking account of all of the evidence, the claimant reasonably should have understood that the representation was not part of the contract.

As for which rule is better, this comes down to one's views on human litigiousness, the fallibility of judges, the institutional considerations that justify the strong rule, and, at bottom, whether one cares more about fairness in particular cases or generally reducing the cost of contracting. I will not try to resolve this conundrum. My interest in this part is more prosaic. It is with how two bodies of law that provide redress for inadvertent misrepresentation outside of contract law deal with the same problem. One is a claim for rescission and restitution on the basis of innocent misrepresentation. The other is the tort action for negligent misstatement. It should be unsurprising that they generally solve the problem the same way. What is surprising is the amount of confusion the problem has created, particularly in the law of negligent misstatement. The handling of inadvertent misrepresentation as problems of tort and equity, and not as problems of contract, is the source of this unnecessary confusion.

A. Innocent Misrepresentation

The doctrine of equitable rescission permits a claimant to rescind a contract for misrepresentation, if the claimant relied on a misrepresentation in entering into the contract and the actor made the misrepresentation for the purpose of inducing the claimant to enter into the contract.[47] Rescission is available even if it was reasonable for the actor to believe that representation was accurate. Restitution also is available unless intervening circumstances make this impractical or unfair to either party or to a third party. In addition, in many American states a damage action is available in

[47] *Hyler v Garner* 548 NW 2d 864 (Iowa 1996) ('the elements of an equitable claim for rescission based on misrepresentation are (1) a representation, (2) falsity, (3) materiality, (4) an intent to induce the other to act or refrain from acting, and (5) justifiable reliance.')

tort for the difference between value given and value received in an exchange generally measured on the date of the exchange.[48] This approximates the value of rescission and restitution on the date of the exchange.

It is fairly well established that an exculpatory term will bar equitable rescission and restitution if it would bar a contract claim on a representation.[49] Some courts reach this result by applying the parol evidence rule to preclude proof of a representation.[50] Some hold that an exculpatory term negates reliance.[51] The logic of this position seems impeccable. The reasons that justify enforcing an exculpatory agreement to shield a party from liability for contract damages when a representation turns out to be inaccurate as well justify holding the claimant to a bad bargain by denying rescission and restitution. If the representation appears in the contract, then the reason presumably is that the contract allocated the risk of the inaccuracy of the representation to the claimant. If the representation is extra-contractual, then there are also evidentiary and institutional reasons.

[48] *Restatement (Second) of Torts* (1977) § 552C so provides. The decision to include innocent misrepresentation in the *Restatement (Second) of Torts* was controversial at the time. Alfred Hill argued in 'Damages for Innocent Misrepresentation' (1973) 73 *Columbia Law Review* 679, and 'Breach of Contract as a Tort' (1974) 74 *Columbia Law Review* 40, that a claim for damages for innocent misrepresentation is best handled in contract law where the parol evidence rule can be used to filter out weak claims. Hill also thought the damage claim too similar to a breach of warranty claim.

FV Harper, F James and OS Gray, *The Law of Torts*, 2nd edn (Boston, Little, Brown and Co, 1986) § 7.7, respond to Hill. They argue that explicit recognition of innocent misrepresentation as a basis for reversing a transaction makes the law more coherent and clear as the same result had been achieved through a 'confusing patchwork' of mostly 'procedural vehicles'. They also identify a practical reason for imposing strict liability for misrepresentation inducing a contract: the 'misrepresenter stands to gain at the expense of the other party to the transaction induced by the misrepresentation'. *Ibid*, at 416. More than an instinctive belief that it is wrong to profit from another's misfortune justifies strict liability in these cases. Strict liability for misrepresentation counteracts the incentive to speak carelessly when error benefits a speaker without requiring individualised determinations of fault.

[49] See, eg, *Wilkinson v Carpenter* 554 P 2d 512 (Or 1976) (seller innocently but mistakenly represented roof was in good condition, the contract stated the property was sold 'as is' with no warranties or representations regarding its quality); *Gibson v Capano* 699 A 2d 68 (Conn 1997) (home seller innocently misstated that termites had not been treated with Chlordane while disclosing extent of termite damage he was aware of, buyers knew of termite problem but not its extent, contract disclaimed reliance on representations); *Creamer v Helferstay* 448 A 2d 332 (Md 1982) (holding parol evidence rule precludes rescission based on a misrepresentation that is inconsistent with the contract); *Hoover v Hegewald* 689 P 2d 965 (Or App 1984) (misrepresentation by broker of number of cattle ranch would support and of number of irrigated and irrigable acres, contract disclaimed representations by seller and its agents and stated that the buyers were relying on own investigation).

[50] Conflicting secondary authority on the point is reviewed in *Wilkinson v Carpenter*, *ibid*, which holds that the rule does apply and that disclaimer precludes redress for innocent misrepresentation. *Restatement (Second) of Contracts* (1981) § 196, Comment *b*, is in accord. For the contrary view that a merger provision or disclaimer does not preclude rescission for innocent misrepresentation see *Halpert v Rosenthal* 267 A 2d 730 (RI 1970).

[51] *Gibson v Capano*, above n 49.

Nevertheless, a handful of cases suggest there are a limited number of situations in which rescission is available for a representation when contract damages are not available. *Britton v Parkin* is an example.[52] A real estate broker mistakenly described property as being zoned commercial in a real estate listing, advertisements and signage. The buyer was allowed to rescind the purchase on the basis of innocent misrepresentation even though the standard form purchase agreement had a merger provision and boilerplate stating the sale was subject to zoning restrictions. The buyer was not allowed to recover expectation damages. A case can be made for the result, particularly if you side with Corbin and against Williston, preferring the weak to the strong form of the parol evidence rule. The result is fair as no one thought there was any doubt about the property's zoning. Allowing rescission and restitution, but not expectation damages, prevents the seller from reaping a windfall from a mutual mistake while not giving the buyer a windfall if the price was a bargain had it been zoned as represented. Of course, allowing such claims sacrifices some of the institutional and evidentiary benefits of the strong form of the parol evidence rule. But the unusual facts in these cases (the misrepresentation and reliance are indisputable), the limited remedy, and the obscurity of the cases limits the sacrifice.

B. Negligent Misstatement

American courts have struggled with the question whether an exculpatory term, such as a disclaimer or a merger provision, will preclude a negligence action for a misrepresentation regarding the subject matter of a contract. Some courts avoid the problem by defining the general scope of the tort of negligent misstatement in a way that precludes an action based on a misrepresentation regarding the actor's contract with the claimant.[53] Other courts have adopted a rule that a valid disclaimer or merger provision will

[52] 438 NW 2d 919 (Mich App 1989). *Norton v Poplos* 443 A 2d 1 (Del 1982) is similar. An advertisement and sign mistakenly listed property as zoned 'M-1.' The buyer was allowed to rescind though a form contract had general merger clause and boilerplate stating that property was subject to restrictions of record. See also *Parkhill v Fuselier* 632 P 2d 1132 (Mont 1981) (listing misrepresented that one acre property had community water supply, form contract said buyer was relying on own investigation and disclaimed oral representations); *Lesher v Strid* 996 P 2d 988 (Or App 2000) (seller gave buyer a map incorrectly indicating four acres had water rights, contract had 'as is' clause but also referred to transfer of water rights).

[53] Examples include a rule that a duty of care is owed only when an actor supplies a claimant with information to guide the claimant in a business transaction with another, *National Can Corp v Whittaker Corp* 505 F Supp 147 (ND Ill 1981) (Illinois law); a rule that a duty of care is owed only when an actor is in the business of supplying the information, *Alderson v Rockwell International Corp* 561 NW 2d 34 (Iowa 1997); and the view that the economic loss rule bars any form of a negligence action between the parties to the contract

bar the tort action as well as a contract action.[54] But some courts have likened negligent misstatement to fraud and held that an exculpatory term will not preclude the action. Approaches vary. Some courts hold that a merger provision or general language of disclaimer will not preclude a claim for negligent misstatement.[55] Other courts go further and hold that contract terms that purport to absolve an actor from liability for a representation have no bearing in a negligent misstatement action other than as facts to be considered by the jury in resolving the tort claim.[56] Under New York law a tort action is available for a misstatement regarding a contract notwithstanding an exculpatory term if the parties are

when the claim relates to the contract's subject matter, see *Duquesne Light Co v Westinghouse Electric Corp* 66 F 3d 604 (3d Cir 1995) (Pennsylvania law); *Apollo Group Inc v Avnet Inc* 58 F 3d 477 (9th Cir 1995) (Arizona law); *Pulte Home Corp v Osmose Wood Preserving Inc* 60 F 3d 734 (11th Cir 1995) (Florida law) (the Florida Supreme Court later repudiated this version of the economic loss rule); *Bailey Farms Inc v NOR-AM Chemical Co* 27 F 3d 188 (6th Cir 1994) (Michigan law); *Danforth v Acorn Structures Inc* 608 A 2d 1194 (Del 1992); *Sebago Inc v Beazer East Inc* 18 F Supp 2d 70 (D Mass 1998) (limiting the rule to sales of goods).

[54] *Sound Techniques Inc v Hoffman* 737 NE 2d 920 (Mass App 2000); *Rio Grande Jewelers v Data General Corp* 689 P 2d 1269 (NM 1984); *Stanley v Miro* 540 A 2d 1123 (Me 1988); *Snyder v Lovercheck*, above n 6; *Brogan v Mitchell International Inc* 692 NE 2d 276 (Ill 1998); *Hodgkins v New England Telephone Co* 82 F 3d 1226 (1st Cir 1996) (Maine law); *Vermont Plastics Inc v Brine Inc* 79 F 3d 272 (2d Cir 1996) (Vermont law); *Lowe v AmeriGas Inc* 52 F Supp 2d 349 (D Conn 1999), affirmed 208 F 3d 203 (2d Cir 2000) (Connecticut law).

[55] See *Agristor Leasing v AO Smith Harvestore Products* 869 F 2d 264 (6th Cir 1989) (Tennessee law); *Keller v AO Smith Harvestore Prods Inc* 819 P 2d 69 (Colo 1991); *Greenfield v Heckenbach* 797 A 2d 63 (Md App 2002); *Robinson v Tripco Investment Inc* 21 P 3d 219 (Utah App 2000); *Grube v Daun* 496 NW 2d 106 (Wis 1992). *Brooks v Timberline Tours Inc* 127 F 3d 1273 (10th Cir 1997) (Colorado law) is authority for the corollary proposition that a clear exculpatory term will preclude a negligent misstatement claim. The two leading cases involve representations by a manufacturer in advertising that an innovative grain storage silo would better preserve the nutritional value of silage by limiting air exposure: *Agristor Leasing v AO Smith Harvestore Products*; *Keller v AO Smith Harvestore Prods*. It turned out that the silos significantly reduced the nutritional value of silage by a combination of air exposure and heat. The claimants did not realise the silo was to blame until they suffered a severe loss of milk production and in one case the deaths of several cows. The manufacturer attempted to avoid liability by invoking contract terms that stated the buyer understood that representations in the advertising were not guarantees and that the buyer did not rely on the representations. While other grounds for recovery were available (including fraud) the claimants obtained a jury verdict on a claim of negligent misstatement. The court of appeals affirmed announcing the rule stated in text. *Horn v AO Smith Corp* 50 F 3d 1365, 1368 (7th Cir 1995) collects other cases involving claims against Harverstore. Often claimants recovered on a fraud claim.

[56] Some cases lump a fraud claim with a negligent misstatement claim to hold that liability cannot be avoided by contract. See *Moffatt Enters Inc v Borden Inc* 807 F 2d 1169 (3d Cir 1986) (Pennsylvania law); *Gibb v Citicorp Mortg Inc* 518 NW 2d 910 (Neb 1994); *Gilliland v Elmwood Properties* 391 SE 2d 577 (SC 1990). *Wilburn v Stewart* 794 P 2d 1197 (NM 1990) lumps these theories with innocent misrepresentation, albeit in dicta *Pearson v Simmonds Precision Products Inc* 624 A 2d 1134 (Vt 1993) holds that it is a question for the jury whether the contract was adequate disclosure. *Formento v Encanto Bus Park* 744 P 2d 22 (Ariz App 1987) holds that the parol evidence rule does not apply to a tort claim.

in a 'special relationship', which is treated as a question for the jury whenever reasonable people could differ on the character of the relationship.[57]

The New York position is bizarre. New York is among the states that will enforce an exculpatory term to preclude a fraud claim. New York also takes the Williston approach, enforcing a 'hard' version of the parol evidence rule. It is impossible to square this with allowing a claimant to get around a contractual disclaimer by claiming a representation was negligent and that the parties were in a 'special relationship'. If the alleged misrepresentation does not appear in the contract and is disputed, then the evidentiary and institutional arguments for making an exculpatory term conclusive evidence of the non-existence of a representation or of reliance are no weaker when an actor allegedly was negligent regarding the accuracy of the representation. If a misrepresentation is a term of a contract or is undisputed, and the claimant sues in tort for negligent misstatement to get around a contract term limiting the right to recover or to sue on the representation, there is no obvious reason for disregarding a contract term allocating the risk of the inaccuracy of a representation to the claimant, merely because we add the fact that the actor was negligent with regards to the accuracy of the representation. Generally, a person who wants to ensure the other party to a contract takes reasonable care to ensure the accuracy of a representation made as an inducement to contract will be better off bargaining for strict liability.

The charitable explanation for the New York position is that the court is using negligent misstatement to move away from the 'hard' version of the parol evidence rule and away from its policy of enforcing exculpatory

[57] As a consequence in New York claimants have had atypical success getting past motions for summary judgment with tort claims for negligent misstatement in contract negotiations. See, eg, *Fresh Direct LLC v Blue Martini Software Inc* 776 NYS 2d 301 (NYAD 2004) (assurances by provider of software regarding its capacity); *CooperVision Inc v Intek Integration Technologies Inc* 794 NYS 2d 812 (NY Sup 2005) (software license and service agreement); *Fleet Bank v Pine Knoll Corp* 736 NYS 2d 737 (NYAD 2002) (assurances by agent of lender that additional financing would be approved; jury question whether there was a special relationship); *Grammer v Turits* 706 NYS 2d 453 (NYAD 2000) (broker did not disclose construction on property adjacent to one-month vacation sublease). New York courts sometimes reject the claim by characterising the parties' relationship in terms that make one cringe. See, eg, *Morris v Putnam Berkley Inc* 687 NYS 2d 139 (NYAD 1999) (holding that employee is not in a 'special relationship' with employer). Oregon has avoided these problems while also applying a general rule imposing a duty of care in supplying information when the parties are in a 'special relationship.' Oregon has a per se rule precluding a duty of care in supplying information in 'arms-length negotiations'. *Onita Pacific Corp v Bronson* 843 P 2d 890 (Or 1992). Further, Oregon law defines a 'special relationship' as a fiduciary-like relationship in which an actor undertakes to exercise discretion on the claimant's behalf. And under Oregon law whether there is a special relationship is for the court to resolve. *Conway v Pacific University* 924 P 2d 818 (Or 1996) (rejecting claim by professor denied tenure that university negligently misled him regarding effect of poor student evaluations on tenure prospects).

agreements to preclude fraud claims. The limitation of the negligence claim to 'special relationships', a peculiar feature of New York law, limits this liberalisation of the rules allowing redress on extra-contractual misrepresentations on a plausible dimension. An oddity of employing the doctrine of negligent misstatement for this purpose is that the actor's negligence regarding the accuracy of an extra-contractual representation becomes an issue. When claimants sue on extra-contractual representations the typical issues are whether the representation was made as an inducement to contract and whether the claimants should have understood that they bore the risk of the inaccuracy of the representation. Whether the other party was negligent regarding the accuracy of the representation is beside the point.

A less charitable explanation for the New York position is that the court has fetishised the classification of negligent misstatement as a tort. The tort action for negligent misstatement dates back to the 1920s in the United States[58] and to the 1960s in the Commonwealth.[59] In both the United States and the Commonwealth, the action was characterised as a tort, and not as a claim for negligent performance of a contractual undertaking, because the defendant supplied misleading information to the claimant gratuitously or pursuant to a contract with a third person. Contract was unavailable because of doctrinal impediments to enforcing gratuitous promises or to third party claims, impediments which have since given way in the United States. The obligation underpinning negligent misstatement resembles contract more than it does the tort of negligence (specifically it resembles contract theories of promissory estoppel and third party beneficiary) for a duty of care must be voluntarily undertaken, typically by an actor inviting a claimant to rely on information. I turn to this point now for it is the key to understanding the power of an actor to contract out of liability for negligent misstatement when the misstatement is not an inducement to contract.

III. CONTRACTING OUT OF LIABILITY FOR NEGLIGENT MISSTATMENT

Recall that in *Snyder v Lovercheck* the buyer, Snyder, included a claim against his agent, Hayek, alleging that Hayek was negligent in failing to warn him that the exculpatory terms in the seller's contract absolved the seller from responsibility for the accuracy of the representation that rye infected only 100 of the 1,900 acres purchased. The court held that Hayek

[58] *Ultramares Corp v Touche* 174 NE 441 (NY 1931) and *Glanzer v Shepard* 135 NE 275 (NY 1922) are the primary cases.

[59] *Hedley Byrne & Co v Heller & Partners Ltd* [1964] AC 465 (HL). [*Hedley Byrne*]

was not subject to negligence liability because Snyder could be expected to have read and understood the implications of the exculpatory terms in the seller's contract. It may seem odd that exculpatory terms in a contract between Snyder and the seller could absolve Hayek from negligence liability to Snyder. This part explains why the result is correct. Hayek owed Snyder no duty of care because Hayek never appeared to invite Snyder to rely on him for advice regarding the contract. That Snyder was a businessman who could read and understand the exculpatory terms supports this conclusion.

Generally, when negligence may result in solely pecuniary harm, an actor has a duty of care in supplying information to a recipient if and only if the actor invites the recipient to rely on the information.[60] More precisely, there is a duty of care if the actor reasonably appears to intend to invite the recipient to attach significant importance to the information in making a decision that may result in a loss if the information is inaccurate or misleading. I will call this 'the requirement of invited reliance'. Stephen Perry has called this 'the requirement of an undertaking'.

That a duty generally requires invited reliance distinguishes the tort of negligent misstatement from the general tort of negligence. It is sufficient for a duty of care in the general tort of negligence that an actor's conduct creates a risk of harm to another that the actor could reduce by the exercise of reasonable care. When people drive on a public road they have a duty of care for the simple reason that driving creates a risk of harm to others that a driver can reduce by driving with care. The existence of a duty does not depend on a driver appearing to intend to invite others to rely on the driver being careful. On the other hand, when an actor gives information to a claimant bearing on a prospective business transaction between the claimant and a third person, the actor undertakes a duty of care to ensure the accuracy of the information only if the actor reasonably appears to intend to invite the claimant to rely on the information.[61]

[60] I make the case for this criterion of duty as an accurate description of United States law in MP Gergen, 'The Ambit of Negligence Liability for Pure Economic Loss' (2006) 48 *Arizona Law Review* 749, 755–60. A similar point has been made regarding Commonwealth law in SR Perry, 'Protected Interests and Undertakings in the Law of Negligence' (1992) 42 *University of Toronto Law Journal* 247 and B Feldthusen, *Economic Negligence*, 4th edn (Scarborough, Carswell, 2000) 48–54, 120–21.

[61] The case of *Stewart Title of Idaho v Nampa Land Title Co Inc* 715 P 2d 1000 (Idaho 1986) illustrates this important distinction. An agent of a title company mistakenly told an agent of an escrow company over the phone that defects in title disclosed in a preliminary report had been cleared. The escrow agent released funds on this basis. The court found no liability because of a trade custom of giving written clearance. The defendant's agent could have foreseen that the claimant's agent might rely on the information given over the phone, but given the trade custom she did not reasonably appear to intend to invite the escrow agent to release funds on the basis of the phone call.

It follows from the requirement of invited reliance that information suppliers have the power to determine whether they have a duty of care, and to determine the content and scope of any duty they undertake, by signaling to a recipient whether and how they intend the recipient to be able to rely on the information. Again this is in sharp contrast with the general tort of negligence. A driver on a public road cannot avoid or limit the duty of care by signaling to others that they should not rely on the driver being careful.[62]

An information supplier can avoid undertaking a duty of care by warning the recipient not to rely on the information.[63] A contract that defines the scope of an actor's engagement will absolve the actor from negligence liability for failing to advise a client about matters outside the scope of the engagement even though it might seem reasonable for the actor to speak up. For example, *Nowell v Dawn-Leavitt Agency* [64] holds that an insurance agent employed to obtain a standard homeowner's policy had no duty to advise a client of the need for flood insurance though the need should have been apparent to her but perhaps would not have been to the client.[65] The holding in *Snyder v Lovercheck* that Hayek had no duty to advise Snyder regarding the contract is analogous. A contract that defines the scope of an actor's engagement may even absolve an actor from negligence liability for affirmatively supplying misleading information

[62] Certain categories of actors are held to a lower duty of care because people who come into contact with them should know they cannot be expected to exercise ordinary care. For example, infants and minors are sometimes held to an age-specific duty of care for this reason.

[63] *Kuehn v Stanley* 91 P 3d 346 (Ariz App 2004) (purchaser of property did not justifiably rely on 'short form appraisal' that stated it was intended only for use in financing). *Cf First National Bank of Newton Cty v Sparkmon* 442 SE 2d 804 (1994) (duty 'may be limited by appropriate disclaimers which would alert those not in privity with the supplier of information that they may rely upon it only at their peril'). Often a disclaimer is given effect by holding reliance in the face of a disclaimer to be unjustifiable. See *Quinn v McGraw-Hill Companies Inc* 168 F 3d 331 (7th Cir 1999) (Illinois law) (disclaimer accompanying bond rating that rating was 'not a recommendation to buy, sell, or hold any such Bonds and may be subject to revision or withdrawal at any time' makes claimant's reliance unjustifiable); *Dakota Bank v Eiesland*, 645 NW 2d 177 (Minn App 2002) (disclaimer that financial statements were an unaudited compilation makes claimant's reliance unjustifiable).

[64] 617 P 2d 1164 (Ariz App 1980).

[65] *Carleton v Tortosa* 17 Cal Rptr 2d 734 (Ct App 1993) (holding that real estate broker who did not undertake to supply tax advice is not subject to negligence liability for failing to advise a client that transactions exchanging property could easily be restructured to avoid tax). *Murphy v Kuhn* 682 NE 2d 972 (NY 1997) is a representative case holding that an insurance agent has no duty to advise a client regarding a need for coverage when the agent does not undertake to render such advice. California imposes a duty on an agent to advise a client regarding insurance only when the agent misstates coverage, the client requests specific coverage, or the agent holds themselves out as having expertise in the specific field: *Fitzpatrick v Hayes* 67 Cal Rptr 2d 445 (App 1997). For additional cases see the Annotation 88 ALR 4th 289.

when the information concerns matters outside the scope of the engagement. For example, *Hill v Bache Halsey Stuart Shields* holds that a stockbroker who gave bad investment advice was absolved of liability by a contract that limited the broker's duties to carrying out the client's orders.[66] Auditors routinely define the extent of their duty to ensure the accuracy of a financial statement by the character of their engagement. For example, *First National Bank of Bluefield v Crawford* holds that in undertaking to supply a 'review report' of financial statements an auditor does not undertake a duty to verify the accuracy of the financial information supplied to it.[67]

Courts do not automatically enforce contract terms that try to preclude or limit an actor's duty in supplying information. Usually when a contract term is disregarded this is consistent with a rule that protects invited reliance because a claimant was reasonable in believing an actor intended to invite the claimant's reliance notwithstanding the exculpatory term. *Ryan v Kanne* illustrates this.[68] A prospective buyer of a small company with poorly maintained financial records told an accountant employed to prepare the company's financial statements that the accounts payable were of particular concern to him and to use 'every conceivable means to determine the accounts payable'. The accountant orally assured the buyer the accounts payable would be accurate within $5,000. Nevertheless, the financial statements delivered by the accountant contained prominent disclaimers that they were 'unaudited statements' and that the accountant expressed no opinion on the validity of the financial information reported. These disclaimers were belied by a note in the comments describing specific measures taken to check the validity of accounts payable. It turned out that the accounts payable were grossly overstated and that a competent audit would have discovered this. The court found the accountant undertook a duty to use reasonable care to verify the accuracy of the accounts payable notwithstanding the disclaimers because of his repeated specific assurances to the contrary.

Less clear is when and how an actor may absolve himself of negligence liability through an exculpatory term if the actor invites the claimant to rely on information supplied by the actor. *Hedley Byrne* raises the issue.[69] The claimant asked its bank to make inquiries to determine if a client could be relied upon to pay for advertising which the claimant was placing on behalf of the client. The bank knew the claimant was on the hook to pay for the advertising if the client did not. The bank replied by a letter stating 'without responsibility on the part of the bank or its officials' that the

[66] 790 F 2d 817 (10th Cir 1986).
[67] 386 SE 2d 310 (W Va 1989).
[68] 170 NW 2d 395 (Iowa 1969).
[69] Above n 59.

client was 'good for its ordinary business engagements'. *Hedley Byrne* is important because it is the first English case recognising the possibility of negligence liability for misstatement. However, the court held that the exculpatory language absolved the bank of a duty. Later English cases supply an alternative rationale—while the bank undertook a duty of care in answering the inquiry the exculpatory term absolved the bank from liability for breach of the duty.[70] If the criterion of invited reliance defines when there is a duty of care in supplying information, then the alternative rationale is preferable in *Hedley Byrne* because the bank probably did appear to intend to invite the claimant to rely on its response in deciding whether to continue to place advertising for the client.[71]

Typically when an exculpatory term is enforced the circumstances are similar to those in *Hedley Byrne*. The exculpatory term is clear and reasonable in the circumstances because the actor supplied the information for a fee that is small in relation to the liability risk. For example, while the cases are dated it seems fairly well established that an exculpatory term in Dun & Bradstreet's subscription agreement absolves the rating agency from negligence liability.[72]

An exculpatory term is not enforced if the term is unclear or if absolving the actor from negligence liability would be unreasonable in the circumstances. *Estey v MacKenzie Engineering Inc* illustrates this point.[73] A home buyer hired an engineering firm to inspect a house prior to purchase to

[70] *Smith v Bush* [1990] 1 AC 831 (HL); *Harris v Wyre Forest District Council* [1989] 2 All ER 514 (HL).

[71] United States cases recognise that a bank has a duty of care in supplying a gratuitous credit reference if the bank has reason to know the claimant is considering whether to extend credit to the subject. *Berkline Corp v Bank of Mississippi* 453 So 2d 699 (Miss 1984) is a leading case. See also *MSA Tubular Products Inc v First Bank and Trust Co, Yale, Oklahoma* 869 F 2d 1422 (10th Cir 1989) (Oklahoma law); *Nevada National Bank v Gold Star Meat Co* 514 P 2d 651 (Nev 1973); TC Russler and SH Epstein, 'Disclosure of Customer Information to Third Parties: When is the Bank Liable?' (1994) 111 *Banking Law Journal* 258, 270.

[72] See *Fidelity Leasing Corp v Dun & Bradstreet Inc* 494 F Supp 786 (ED Pa 1980) (Pennsylvania law); *Hong Kong Export Credit Insurance Corp v Dun & Bradstreet* 414 F Supp 153 (SDNY 1975) (New York law); *Xiques v Bradstreet Co* 24 NYS 48 (Sup Ct 1893), affirmed 36 NE 740 (NY 1894). *Duncan v Dun* 8 F Cas 9 (CCED Pa 1879) holds the clause also bars a claim for gross negligence. *Globe Home Improvement Co v Perth Amboy Chamber of Commerce Credit Rating Bureau* 182 A 641 (NJ App 1936) also gives effect to a general disclaimer when information is supplied for a nominal payment. The case holds there is no liability beyond recovery of the amount paid for negligence in supplying a credit report where the report stated the information 'is based upon information obtained in good faith by the agent from sources deemed reliable, the accuracy of which, however, is in no manner guaranteed.' The opinion emphasises the small amount paid for the report. *Gale v Value Line* 640 F Supp 967 (DRI 1986) holds that the statement '[factual] material is obtained from sources believed to be reliable but cannot be guaranteed' in a publication ranking securities is insufficient to preclude liability for negligence while finding no duty because the information was misleading by omission.

[73] 927 P 2d 86 (Or 1996). *Schaffer v Property Evaluations Inc* 854 SW 2d 493 (Mo App 1993) is similar.

identify major structural defects and settling. The contract limited the firm's liability to $200, which was the price paid for the inspection. The buyer ended up incurring a substantial loss because of undiscovered structural defects that a competent inspection would have revealed. The court held the exculpatory term did not shield the firm from liability in excess of $200 on a negligence claim. It declined to establish a per se rule and instead based its decision on the facts that the claimant was a consumer, the exculpatory term did not explicitly cover a negligence claim, the magnitude of the fee paid for the inspection, and the importance of the inspection to the claimant. There also is some authority for a rule that an exculpatory term will not shield an actor from liability for gross negligence.[74]

The lesson to be drawn is that an actor generally has no duty of care in disseminating information that may affect a recipient's dealings with third persons. A duty arises only when an actor invites a recipient to rely on the information. Actors can determine by contract whether they have a duty, and its content and scope, by making apparent when, and to what extent, they invite reliance. But once an actor undertakes a duty it is not so easily disclaimed. The pattern is a familiar one in contract law. People can generally determine when they are under a contractual obligation by avoiding an apparent commitment, but once they make an apparent commitment they may absolve themselves of the attendant obligation only when this is reasonable in the circumstances.

IV. CONCLUSION

The law of contract has a fairly well-defined relationship to the law of deceit because the practice of deceit is inimical to the practice of contract. The possibility of deceit undermines trust, which is essential to contract. Deceit and contract are so adverse that we are deeply sceptical as to whether a person ever would intelligently agree to bear the risk of being deceived about a contract. The best argument for enforcing exculpatory agreements to preclude fraud claims is that the agreements shield honest people from unfounded claims of fraud. At bottom, the argument for enforcing exculpatory agreements grounds on a deeply pessimistic view of human litigiousness, and of the capacity of courts to distinguish good fraud claims from bad, and on a very optimistic view of private ordering. Even if one is sympathetic to these views, one should consider the possibility that rules requiring that fraud be pled with specificity, and

[74] This exception is stated in the cases cited above n 72 enforcing the exculpatory term in Dun & Bradstreet's subscription agreement.

proven by clear and convincing evidence, provide sufficient protection against unfounded claims of fraud while at the same time enabling courts to redress proven fraud.

The law of contract has a less well-defined relationship to bodies of law outside of contract that provide redress for inadvertent misrepresentation. This is largely attributable to the obscurity of the law of equitable rescission and restitution and the immaturity of the law of negligent misstatement. I have argued that principles of contract law generally should control on these issues. Contract law rules that protect against unfounded claims of extra-contractual representations—this is the work of the parol evidence rule in the United States—should also protect against claims in equity, restitution, or tort. And contract terms that allocate the risk of the inaccuracy of a representation to a claimant should also apply to claims in equity, restitution, and tort. Often in a negligent misstatement case an actor's obligation to a claimant is not based on a contract because the actor provides information to a claimant gratuitously or pursuant to a contract with another. In such cases the requirement for a duty of care of invited reliance generally enables an actor to determine the existence and define the scope of the duty to a claimant.

11

Assignments, Trusts, Property and Obligations

ANDREW TETTENBORN

I. INTRODUCTION

DESPITE THE FACT that two sizeable books have recently been written about the assignment of contractual rights,[1] the topic is not something that overly excites contracts scholars. It is nevertheless a very important topic in practice,[2] and also one that raises some awkward questions about the nature of the rights arising under an assignment, not to mention issues about the law of trusts and the disputed territory lying between property and obligation.

Two English cases have recently thrown some of these points into sharp relief, and for that reason seem worth at least a brief article. The cases both involve what looks like a narrow question, namely, how far (if at all) a third party may use processes similar, but not identical, to assignment in order to enforce for his or her own benefit a contractual right otherwise non-assignable.

In the first case, *Don King Productions Inc v Warren*,[3] two boxing promoters with substantial pugilistic portfolios went into partnership. They agreed at the time, and confirmed later, that the benefit of all their contracts with individual boxers should be joint assets, held on trust for

[1] M Smith, *Law of Assignment: The Creation and Transfer of Choses in Action* (Oxford, Oxford University Press, 2007) and G Tolhurst, *The Assignment of Contractual Rights* (Oxford, Hart Publishing, 2006).

[2] In England and Canada, one need only mention in this connection the vital role assignment plays in lending against receivables, in securitisation and in the working out of multiparty construction contracts. In the American context, one can add the extensive use of assignment, especially of malpractice claims, in settling litigation, on which see, eg, R Walters, 'The Unwitting Attorney, the Desperate Client, and the Perpetuation of the New York Power Play: a Proposal to Ban Voluntary Assignments of Legal Malpractice Claims via New York General Obligations Law Section 12–101' (2005) 3 *Cardozo Public Law, Policy and Ethics Journal* 543.

[3] [2000] Ch 291 (CA), aff'g [2000] Ch 295 [*Don King*].

the partnership. The arrangement, like many similar ones, subsequently soured, and the partnership fell to be dissolved. At this point the issue arose as to whether the rights under the contracts did indeed form part of its assets. The partner who had brought in the more valuable contracts raised an ingenious argument that they did not. The agreements, he pointed out, depended fundamentally on the skill and flair of the promoter appointed under them, and were for that reason in their nature too personal to be assigned.[4] For good measure most of them also contained express anti-assignment clauses, which were indisputably valid.[5] It must follow, he argued, that the right to enforce them could not be vested in anyone other than the original contractor.

Lightman J disagreed with these contentions. One of his reasons was relatively uncontroversial, namely that however unassignable an obligation might be in principle, there was no difficulty in enforcing a contractual agreement between partners to treat the agreement as if it was partnership property for the purpose of deciding who got what when the arrangement collapsed. This seems plainly correct, and will not be discussed further. Lightman J's second reason, by contrast, is much more difficult. Even if a contract was of such a personal nature as to be unassignable, or contained a clear contractual prohibition on assignment, according to Lightman J this did not mean that there was any bar as a matter of law to the promisee holding the benefit of the contract on trust for a third party. It might be true that allowing such a trust de facto emasculated the principle that personal agreements could not be assigned and gave potential quasi-assignees an end-run around otherwise effective anti-assignment clauses, since a trust of an obligation which bound the obligee to enforce it as the beneficiary's catspaw has a result to all intents and purposes indistinguishable from that of an equitable assignment to that same beneficiary. Nevertheless, this was not objectionable as such, at least in the absence of a specific provision forbidding not only assignment but the creation of trusts. On appeal, the Court of Appeal upheld Lightman J's reasoning, though Morritt LJ added, somewhat enigmatically, that although an otherwise inalienable obligation might be held in trust, this would not as such give the beneficiary any right to interfere in its enforcement.[6]

[4] Under the principle in *Nokes v Doncaster Amalgamated Collieries Ltd* [1940] AC 1014 (HL).

[5] Any lingering doubts on this were dispelled by the House of Lords in *Linden Gardens Trust Ltd v Lenesta Sludge Disposal Ltd* [1994] 1 AC 85 [*Linden Gardens*]. The law in Canada is similar: *Brio Beverages (BC) Inc v Koala Beverages Ltd* [1999] 6 WWR 219 (BCCA).

[6] See *Don King*, above n 3, at 335–6.

Don King was followed, in a relatively uncontroversial way, in two subsequent English decisions.[7] But neither of these directly involved the question of allowing trust beneficiaries to enforce unassignable contracts. This question did arise bluntly, however, when the Court of Appeal was faced with *Barbados Trust Co v Bank of Zambia*.[8] Briefly, Bank of America (BOA) held Zambian government bonds, expressed by their terms to be assignable to banks but not to other financial institutions. BOA nevertheless purported to assign these bonds to a non-bank, and following a series of further transfers they ultimately ended up in the hands of Barbados Trust Co (BTC), another non-bank, which wanted to sue on them. The restrictions on assignment in the bonds obviously amounted to a major roadblock, meaning that the right to sue remained all along in BOA, but, prompted by BTC, BOA sought to repair this problem by declaring itself trustee of the bonds for BTC. BTC then sued the Bank of Zambia wearing its hat as trust beneficiary instead of assignee, following the standard practice of joining BOA, the trustee, as nominal defendant.[9] Bank of Zambia for its part argued that the declaration of trust was a mere subterfuge to circumvent the contractual limitations on assignment, and that BTC should remain unable to obtain indirectly what it could not get directly. BTC did indeed fail to recover, on the ground that BOA's own title had been defective (for reasons not relevant here) and that hence BTC's fell with it. However, a majority of the Court of Appeal also decided, reversing Langley J and explicitly endorsing the reasoning in *Don King*, that the bar on assignment to non-banks was an irrelevance.[10] A declaration of trust, the court held, was not the same thing as an assignment: the prohibition on assignment did not in its terms prohibit a declaration of trust, and whatever the position might have been had it purported to do so, that meant that BTC's title to sue as trust beneficiary was beyond attack. In contrast, Hooper LJ dissented, making the obvious contrary point. Whatever analytical differences might exist between assignments and declarations of trust, he observed, the result of allowing BTC to sue as the beneficiary of a trust of a debt was precisely the same as allowing it to sue

[7] See *Swift v Dairywise Farms Ltd* [2000] 1 All ER 320 (CA), where a trust was pressed into service to allow hypothecation to a lender of proceeds of an otherwise nontransferable milk quota arising under European Union agricultural marketing legislation; *John Taylors (A Firm) v Masons (A Firm)* [2001] EWCA Civ 2106, which applied the rule in *Keech v Sandford* (1726) Sel Cas T King 61, 25 ER 223 (Ch) to renewals of unassignable licenses.

[8] [2007] EWCA Civ 148, [2007] 1 Lloyd's Rep 495 (CA) [*Barbados Trust*].

[9] The practice was standard in that it was implicitly approved in *Les Affréteurs Réunis SA v Leopold Walford (London) Ltd* [1919] AC 801 (HL). See also *Performing Right Society Ltd v London Theatre of Varieties Ltd* [1924] AC 1 (HL); *Vandepitte v Preferred Accident Insurance Corporation of New York* [1933] AC 70 (PC) 74 [*Vandepitte*].

[10] Waller and Rix LJJ.

as the equitable assignee of that debt, and it was a curious rule of law that, having driven this remedy out of the front door, promptly let it in by the back door.

In this article I argue that *Don King* and *Barbados Trust* not only reach a woefully unrealistic commercial result but more importantly misunderstand the nature of equitable assignment. In particular, the supposed contrast between trusts and equitable assignments, I suggest, is a false one.

II. EQUITABLE ASSIGNMENT: THE PECULIAR ENGLISH APPROACH[11]

If you ask a civil lawyer, or for that matter a Scottish jurist, what an assignment is, the answer is relatively simple: it is a wholesale substitution of creditors.[12] A creditor C with a right to exact performance from a debtor D has voluntarily transferred that right to an assignee A, who is thereby substituted for C and able to pursue D.

Unfortunately the common lawyers, for reasons never adequately explained, could not reconcile themselves to this idea that an obligation contracted in favour of C could morph into one owed to A, even if C was happy for it to do so.[13] Whatever C might intend, in that respect he or she remained the creditor and the only person entitled to claim performance. This was why, given the commercial necessity of allowing assignment under some guise or other, equity had to be roped in to fill the gap. Unfortunately, it did so by using two, or possibly three, disjointed approaches simultaneously.[14]

To begin with, equity recognised what today we call assignment proper. If C entered into some agreement with A envisaging the transfer from C to A of the benefit of C's rights against D, equity intervened from the seventeenth century to ensure that the right was enforced,[15] and that when

[11] We are not, for obvious reasons, concerned here with statutory assignments under the provisions of s 136 of the Law of Property Act 1925 (UK) 15 & 16 Geo V c 20 or its Canadian analogues (eg Law and Equity Act RSBC 1996 c 253, s 36).

[12] On French law see the bald statement in B Starck *et al*, *Droit Civil: les Obligations*, 6th edn (Paris, Litec, 1998) vol 2 § 1591 (*cessionaire de ces créances 'pourra s'en prévaloir'*). See also J Ghestin, *Traité de Droit Civil: La Formation du Contrat*, 3rd edn (Paris, Librairie Générale de Droit et de Jurisprudence, 1993) § 733. For Scotland, see D Walker, *Principles of Scottish Private Law*, 2nd edn (Oxford, Clarendon Press, 1975) vol 2, 1750 (the effect of assignation is 'to divest the cedent completely and to put the assignee in his place, entitled to sue for the enforcement of the assigned right, … to receive payment and grant a good discharge therefore'). The third edition of the same book, perhaps oddly, seems to omit this lapidary description.

[13] On which see W Cook, 'The Alienability of Choses in Action' (1916) 29 *Harvard Law Review* 816; W Holdsworth, *History of English Law* (London, Methuen & Co Ltd, 1925) vol 7, 520–25; Smith, above n 1, at §§ 5.04–5.09. The chief grounds for complaint seem to have been the inherently personal nature of personal actions and a loosely-articulated fear of the buying up of proceedings by the unscrupulous and litigious.

[14] What follows is neatly explained in Smith, above n 1, at §§ 6.02 ff.

[15] Holdsworth, above n 13, at 519.

it was, A got the benefit of it. This was accomplished by allowing A to sue in equity to force C to invoke his or her claim, and then adding a congeries of ancillary equitable rights to protect A's position in other ways.[16] After fusion,[17] A sued and joined C in a single proceeding, with the ancillary rights directly protected.[18]

Secondly, there was a parallel development, though one with much the same long-term result. Lawyers had long talked about a debt or contractual right being an item of property:[19] in other words, an asset, and hence something that might figure in the 'things' section of a putative civil code as well as under the 'obligations' rubric. As a result, there had logically to be a role for the law of trusts proper. If a piece of land could be held on trust, with legal property in one person and beneficial ownership in another, the same must follow for a debt or other contractual right: it could be owned in equity by someone other than the person appearing to be the obligee.[20] It followed that a contractual right could also be assigned de facto by the simple mode of C declaring himself or herself a bare trustee of it for A.[21]

We are primarily concerned with assignments and declarations of trust. However, it is worth adding for the sake of completeness a third possible form of transfer arising from the incorporation into the law of assignment of the maxim that equity regards as done that which ought to be done. It became clear by the end of the nineteenth century that a contract for value by C to transfer a right—whether present or future—to A caused the right in question to stand transferred in equity either immediately or as and when it arose.[22] This rapidly became the most significant form of assignment, as witnessed in the growth of debt factoring and general charges over book-debts, both of which depend on it. But whether a promise to assign is regarded as equivalent to an assignment proper of the right, or a

[16] For example, by enjoining D from pleading a set-off incurred, or a release arranged with C, after D knew of A's position as assignee. See R Derham, *Set-off*, 3rd edn (Oxford, Oxford University Press, 2003) §§ 17.02–17.03.

[17] Or rather, in the English context, from the time of the partial fusion effected by the Common Law Procedure Act 1854 (UK) 17 & 18 Vict c 125, allowing common law courts to give effect to equitable interests.

[18] For example, if D had a set-off against C incurred after notice of the assignment, this would simply be ignored and judgment given for A.

[19] See, eg, *Fitzroy v Cave* [1905] 2 KB 364 (CA) 372–3 (Cozens-Hardy LJ); *Ellis v Torrington* [1920] 1 KB 399 (CA) 411 (Scrutton LJ); Tolhurst, above n 1, at § 3.20.

[20] For a straightforward expression of this idea see, eg, *Fletcher v Fletcher* (1844) 4 Hare 67, 67 ER 564 (ChD) 74 (Wigram V-C).

[21] Smith, above n 1, at §§ 6.33 ff. It is also possible, of course, for a trust of a contractual right to arise out of some more complex trust relation, where the trustee has active duties. A straightforward example is a trustee for bondholders. But this article is concerned with the trust as a simple means of *de facto* assignment: ie a bare trust and no more.

[22] As in the rule in *Holroyd v Marshall* (1862) 10 HL Cas 191, 11 ER 999. Its extension to debts generally was cemented by *Tailby v Official Receiver* (1888) 13 AC 523 (HL).

bare trust of it (if there was a difference between the two) has never been made clear.[23] One suspects that this is because the point does not often matter in practice. In either case A is entitled in equity to have the right exercised for his or her benefit and not C's, even if, for example, C is insolvent, and this is normally all that matters.

III. TRUSTS AND EQUITABLE ASSIGNMENTS: TWO LEGAL RESULTS OR ONE?

Having dealt with the theory of equitable assignments, we can now turn to assignment and the creation of a bare trust in more detail. As described above, these are clearly two different processes. One simply requires an agreement, however expressed, to transfer the benefit of a contractual or other right to someone else.[24] The other involves meeting all the requirements—certainty of intention, certainty of subject matter, and so on—necessary for the creation of a valid express trust,[25] which is not the same thing at all. But this is not a point that matters here. The important question is not whether assignments and trusts are generated in different ways, but rather whether they also give rise to sufficiently different legal results to justify the law's discriminating between them, as *Don King* and *Barbados Trust* clearly suggest.[26] It is my suggestion that, if we compare them properly, they do not.

We begin by looking at the effects of an equitable assignment proper. These, it is suggested, can be effectively summed up in five propositions. First, A prevails, like any other valid transferee, against the creditors of C if C is insolvent.[27] Second, the best explanation as to why A prevails is that from the moment of the equitable assignment C is treated by equity as holding the benefit of the obligation assigned on constructive trust for A as

[23] But the latter is perhaps more likely. In particular, Lord Fitzgerald in *Tailby v Official Receiver*, *ibid*, at 546 based the rule on the idea that 'whenever persons agree concerning any particular subject, that, in a Court of Equity, as against the party himself, and any claiming under him, voluntarily or with notice, *raises a trust*' (emphasis added).

[24] See, eg, *William Brandt's Sons & Co v Dunlop Rubber Company Ltd* [1905] AC 454 (HL) 462 (Lord M'Naghten).

[25] On which see, eg, J McGhee (ed), *Snell's Equity*, 31st edn (London, Thomson, 2005) §§ 20–15.

[26] An analogy may make this clearer. At common law a gift of a chattel may be made in two entirely different ways: by delivering it or alternatively by executing a deed of gift. Nevertheless, the legal result of these two different actions is precisely the same: the donee becomes owner of the chattel concerned and gets all the rights one would expect an owner to have.

[27] The assignee is, of course, subject to specific limitations, such as the insolvency rules providing for the annihilation of transfers in fraud of creditors. But so is the transferee of any other asset, so this changes nothing.

beneficiary.[28] Third, A can insist—originally by separate proceedings for a common injunction, but since the fusion of law and equity by proceedings directly against D,[29] with C joined as a formal party[30]—that the right be exercised for his or her benefit, and hence take control of its enforcement. Fourth, A takes 'subject to equities'. More precisely, this means (i) that A can only enforce the right 'warts and all', that is, subject to any defences or limitations[31] and (ii) that A will be subject to certain rights of set-off—broadly, all connected counterclaims[32] and certain unconnected ones arising before notice of the assignment.[33] Fifth, A has the right to protection from any post-notice set-off arising between assignor and debtor[34] and any post-notice agreement between assignor and debtor to release or change the contractual rights involved.[35]

Now, how far do these legal results apply to a bare express trust of a debt? Let us take each in turn.

A. Prevalence in Insolvency

The interests of any trust beneficiary prevail against the creditors of an insolvent trustee, and choses in action are no exception. We do not need to take this point any further.

[28] Smith, above n 1, at § 6.12. But *cf* J Story, *Commentaries on Equity Jurisprudence*, 3rd English edn (London, Sweet & Maxwell, 1920) 432, which regards an assignee as equivalent to the beneficiary of an express trust.

[29] Or, more accurately in the English context, since the Common Law Procedure Act 1854, above n 17, allowed common law courts to give direct effect to equitable rights.

[30] *Durham Brothers v Robertson* [1898] 1 QB 765 (CA) 770 (Chitty LJ); *Performing Right Society Ltd v London Theatre of Varieties Ltd* [1924] AC 1 (HL) 19 (Viscount Finlay). The fact that the original creditor must be joined in the action is a side-issue here, being on the modern view largely a matter of procedure: eg *The Aiolos* [1983] 2 Lloyd's Rep 25 (CA) 32, 34; *Weddell v Pearce & Major* [1988] 1 Ch 26. See also n 9 above.

[31] See, eg, *Athenaeum Life Assurance Co v Pooley* (1853) 3 De G & J 294, 44 ER 1281 (CA); *Graham v Johnson* (1869) LR 8 Eq 36.

[32] *Newfoundland Government v Newfoundland Railway Co* (1888) LR 13 AC 199 (PC). More precisely, the opposable claims are those that between D and C would amount to equitable or 'transactional' set-off.

[33] *Watson v Mid Wales Railway Co* (1867) LR 2 CP 593. More precisely, this covers claims otherwise covered by the principles contained in the Statutes of Set-off 1729 and 1735 (2 Geo II c 22 and 8 Geo II c 24 respectively). The reason for limiting this susceptibility to pre-notice cross-claims is that a debtor who gives credit to the creditor not knowing of an assignment should not lose any rights of set-off he or she would otherwise get, but a debtor who knows of the assignment deserves no such indulgence.

[34] *Ibid*.

[35] *Brice v Bannister* (1878) 3 QBD 569 (CA).

B. The Nature of the Beneficiary's Interest in the Obligation Concerned

As mentioned above, the interest of an equitable assignee, on the best explanation, arises from a constructive trust.[36] That of a trust beneficiary proper arises, *ex hypothesi*, from an express trust. There is clearly a difference here. Nevertheless, it is a difference of source rather than result. In both cases the assignee or beneficiary ends up as the equitable owner of the obligation, and in both cases can enforce the obligation in the same way—namely by the simple expedient of joining the assignor or trustee.

C. Control over Enforcement

Logically, this feature—the ability of the beneficiary to control and if necessary to compel enforcement—must apply in the trust scenario as much as in the case of assignment, at least as concerns a bare trust (it may be different where the terms of a trust leave some discretion to the trustees as to how to handle the trust assets: we will return to this later). This certainly seems to follow from the cases on trusts of a promise generally, such as *Vandepitte v Preferred Accident Insurance Corporation of New York*.[37] There the Privy Council implicitly approved[38] the procedure whereby, assuming the benefit of a contractual obligation is indeed held on trust,[39] proceedings can be brought by the beneficiary in his or her own name, with the trustee joined as co-plaintiff or co-defendant as required. This procedure, moreover, was approved in *Barbados Trust* itself.[40] Moreover, this makes good sense. For a bare trustee of a contractual right to refuse point-blank to enforce it when asked to do so by the beneficiary[41]—and hence in effect to nullify it—would seem to be as blatant a breach of trust as one could imagine. If so, it seems fairly obvious that the beneficiary ought to be able to prevent such a breach of trust by some device such as that described. The choice, in other words, as to whether the obligation is enforced lies squarely with the plaintiff.

The only possible counter-argument comes in a slightly curious throw-away suggestion by Lightman J in *Don King* that the beneficiary under a trust might not in fact be able to control enforcement.[42] This, he reasoned, was because the rule in *Re Brockbank* said that a beneficiary could not

[36] See above, text accompanying n 28.
[37] Above n 9.
[38] *Ibid*, at 79 (Lord Wright).
[39] In fact it was not so held in *Vandepitte*, but that does not alter the point being made here.
[40] See in particular Rix LJ in *Barbados Trust*, above n 8, at [99].
[41] Assuming adequate arrangements made for costs and the like.
[42] Above n 3, at 321.

dictate to a trustee how to exercise his functions.[43] But this must be misconceived, for two reasons. First, the point of *Re Brockbank* is not that a beneficiary cannot insist on a trustee observing the terms of the trust, but rather that a beneficiary, even if solely entitled, cannot tell the trustee to commit an act that would amount to a breach of trust.[44] If there is no discretion given to the trustee—and it is suggested that a bare trust of an obligation cannot leave it up to the trustee to decide whether to enforce it—there is no reason for the rule to apply.[45] Second, if Lightman J is right, and the beneficiary of a trust of a chose in action cannot insist that the trustee actually exercise it, this seems almost entirely to defeat the point of the trust in the first place. Put shortly, if the beneficiary does not have this right, what rights worth having does he possess?[46]

D. 'Subject to Equities'

What of the fourth principle, the 'subject to equities' rule? One aspect of this is easy: the rule that the benefit of an obligation can only be taken warts and all must apply in the law of trusts. Although there appears to be no direct English authority on the application of this rule to trust beneficiaries (as against assignees),[47] it must be the case that the beneficiary of a trust of a right gets the right to sue that the trustee has, no more and no less. You cannot declare yourself trustee of something you do not have. It follows that in so far as an obligor has an excuse for refusing performance, such as an ability to plead non-liability due to fraud or error, or breach of fiduciary duty, this right must remain.[48]

What of the other aspect of the 'subject to equities' rule—counterclaims available to the debtor? In so far as connected counterclaims are concerned (such as those arising out of equitable set-offs which would have been

[43] [1948] Ch 206.

[44] In that case, the beneficiaries had purported to tell a surviving trustee whom to appoint as a new trustee, something which was specifically left to the trustee's own discretion.

[45] *Cf Citibank NA v MBIA Assurance SA* [2007] 1 All ER (Comm) 475, where the Court of Appeal certainly assumed that a trustee may, if the terms of the trust so provide, follow the directions of a beneficiary in exercising rights against third parties.

[46] See Millett LJ in *Armitage v Nurse* [1998] Ch 241 (CA) 253: 'there is an irreducible core of obligations owed by the trustees to the beneficiaries and enforceable by them which is fundamental to the concept of a trust. If the beneficiaries have no rights enforceable against the trustees there are no trusts.'

[47] But there is authority concerning those entitled to the benefit of contractual rights by subrogation: see *The Front Comor* [2005] 2 All ER (Comm) 240 (QB) and *The Jay Bola* [1997] 2 Lloyd's Rep 279 (CA) holding that if a right can only be exercised by the original contractor subject to an arbitration clause, the same goes for an insurer claiming by subrogation. And if a beneficiary takes subject to certain counterclaims (see *Murphy v Zamonex Pty Ltd* (1993) 31 NSWLR 439 (SC)) then it must follow a fortiori that he takes subject to defences.

[48] This also seems to be implicit in *Murphy v Zamonex Pty Ltd*, *ibid*.

available as between debtor and trustee), logic dictates that they must be opposable to a trust beneficiary as much as to an assignee. To begin with, in so far as they would have amounted to an equitable defence against the creditor, the same reasoning applies as to the other defences mentioned above. In any case, the rationale for allowing such a set-off is that the connection between the right being sued on and the debtor's cross-claim is so close that it would be inequitable to the debtor to determine the former without the latter. If this is right, it should make no difference that the person seeking to enforce the right is acting as trustee for someone else who actually stands to benefit. The injustice to the debtor is, after all, precisely the same in both cases. Indeed, this has been held to be the case both in Australia[49] and in England.[50]

What about unconnected counterclaims against the trustee? In the assignment context, these are (as mentioned above) available against an assignee if created before the debtor knew of the assignment, on the basis that they would have afforded a defence under the Statutes of Set-off had the assignor sued, and that an assignment of which the debtor was unaware at the relevant time should not alter the position. There seems to be no authority deciding whether the same thing applies to a trust of the obligation, but the reasoning in the assignment cases must apply here too.[51] A debtor who thinks he or she is obtaining a set-off ought to be no more affected by an unknown trust than by an unknown assignment.

E. Insulation from Post-Notice Equities

The question as to whether a beneficiary under a trust is insulated from post-notice cross-claims or changes is a little more awkward. In the context of the assignee, what is now the established rule arose in the days before fusion from the Court of Chancery's practice of supporting the assignee's right with certain ancillary orders. These notably included injunctions restraining the debtor from relying, in any action by the creditor, on either the Statutes of Set-off (in the case of a cross-claim otherwise covered by

[49] *Murphy v Zamonex Pty Ltd*, *ibid*. The case involved a claim on a loan by a trustee suing as such. The court held that a trade practices claim by the borrower against the trustee's predecessor could be pleaded as equitable set-off. See also *Doherty v Murphy* [1996] 2 VR 553 (SC).

[50] *Penwith District Council v VP Developments Ltd* [2005] EWHC 259 (Ch) (Laddie J). An unsatisfied arbitration costs order had been made against a construction company. The company then claimed in related arbitration to recover alleged underpayments. This claim was available for set-off so as to prevent a winding-up on the basis of the costs order, irrespective of the fact that the company was insolvent and suing entirely as trustee for benefit of its creditors.

[51] And indeed it is assumed to do so in Derham, above n 16, at § 17.96, where it is stated (though without authority) that the rules of statutory set-off are the same for obligations held on trust as for assigned ones.

them) or, as the case might be, on any agreement to release or reduce the rights of the original creditor.[52] Since fusion, life is more straightforward: these rights of the assignee are directly recognised by the courts and the set-offs that would otherwise be available to the debtor are ignored. But can the same reasoning be applied to rights simply held on trust?

At first sight, it might indeed seem not, thus providing at least some clear blue water between trusts and assignments. Consider a common kind of trust, such as a trust of a landed estate or a business, or a commercial trust supporting an issue of bonds whereby one custodian holds the legal interest with the actual investors as beneficiaries. No one would seriously argue that a debtor to such a trust, even if he knew he was dealing with a trustee, should be unable to deal with, and if necessary compromise, the trust's claims against him without the express permission of every beneficiary. If this were the law, it would negative the whole point of such arrangements, which is to leave the day-to-day management of the business or portfolio to the trustees' informed discretion. Moreover, it can also be pointed out that in at least one arguably analogous situation, the beneficiary is not protected. Whereas an assignment plus notice to the debtor locks the assignee's rights in against subsequent alterations, the process of insurer's subrogation to claims of the assured does not. On the contrary, it is well-established that even where an insurer does have subrogation rights over the assured's claim against the obligor, those rights can be validly released or altered by the assured, and if they are, the insurer's only remedy is against the assured.[53]

Whether these points will hold water where we are dealing with a bare trust of an obligation, however, is highly doubtful.

As for the point about a trust of a business or a bondholders' trust, the response is that not all trusts are the same. The functions of trustees, and with them the duties they owe to beneficial owners, vary. In particular, there is every difference between the kinds of trust mentioned in the previous paragraph and the kind of bare trust that does duty as a surrogate assignment. Complex trusts by their nature—and not infrequently by express stipulation—give the trustee an active role in managing the trust property, and the discretion that goes with that role.[54] A bare trust, by contrast, is just that: rather like a well-trained footman, the trustee has little or no function besides holding the legal interest for the benefit of the *cestui que trust*. And if the trustee has no discretion or real independent function, then third parties dealing with the trustee have no reason to expect him or her to be able to release the beneficiary's rights without the latter's consent.

[52] See above, n 16.

[53] *West of England Fire Insurance Co v Isaacs* [1897] 1 QB 226 (CA).

[54] Even on occasion to the extent of disregarding the express wishes of the beneficiary: see *Re Brockbank*, above n 43.

Furthermore, it is suggested that the analogy between bare trusts and the insurer's right of subrogation is false. First, the assured retains a real interest on his or her own behalf in the claim even where there is subrogation: there may, for example, be uninsured losses, or the possibility of recovery over and above the amount received from the insurer. Insurance subrogation is therefore far removed from the kind of bare trust we are considering here, whose object is simply to transfer the benefit of the obligation lock, stock and barrel to the beneficiary. Second, it is also rather dubious whether subrogation creates a true trust of the obligation in favour of the insurer in any case. Although the matter is not beyond argument, the better position seems to be that the insurer gets a proprietary right over the proceeds of the right, but not over the right itself.[55] If so, it is not surprising that the right itself, remaining beneficially owned by the assured, can be altered by him or her at will.

Admittedly, in the absence of clear pre-fusion authority on what the Court of Chancery would have done about a plea of set-off or release in the case of a bare trust, it is difficult to be absolutely certain as to what the law is. Nevertheless, what little authority there is in England suggests that where an obligation is, to the knowledge of a third party, held on a bare trust, then the beneficiary is indeed protected from alterations to his or her detriment.[56]

IV. MORE ON TRUSTS AND ASSIGNMENTS

Pulling together the above strands, it is difficult to see any difference of importance between an assignment and a trust. The only distinction found, that one creates a constructive trust and the other an express trust, may be technically genuine, but its rational significance is nil.[57] If this is correct, then the major plank in the recent decisions on assignment—that there is

[55] *Re Ballast plc* [2007] BCC 620 (Ch). See also *Morris v Ford Motor Co Ltd* [1973] QB 792 (CA) 800, where Lord Denning MR doubted if subrogation was anything like assignment.

[56] See Jessel MR in *Re Empress Engineering Co* (1880) 16 ChD 125 (CA) 129: 'A mere agreement between A. and B. that B. shall pay C. (an agreement to which C. is not a party either directly or indirectly) will not prevent A. and B. from coming to a new agreement the next day releasing the old one. *If C. were a cestui que trust it would have that effect*' (emphasis added). See also Smith, above n 1, at § 13.11 and *cf* AW Scott, *Scott on Trusts*, 3rd edn (Boston, Little, Brown & Co, 1967) vol 4, 2517 to the same effect. There is one apparently contrary authority, *Gibson v Winter* (1833) 5 B & Ad 96, 110 ER 728 (KB) but this was doubted by Lord Campbell in *De Pothonier v De Mattos* (1858) El Bl & El 461, 120 ER 581 (KB) 483 and by the Supreme Court of Canada in *Culina v Giuliani* [1972] SCR 343, 22 DLR (3d) 210.

[57] It is true that Waller LJ in *Barbados Trust*, above n 8, at [43] also observed another difference, namely that an equitable assignment, unlike a trust, might be convertible into a statutory assignment under s 136 of the Law of Property Act 1925, above n 11, by the giving of a suitable written notice. This is correct. But can this matter?

some special feature of equitable assignment that makes it different from other equitable interests—disappears.

Before we conclude that *Don King* and *Barbados Trust* were wrong to allow trust beneficiaries to subvert the rule against assignment of non-assignable rights, however, there is one more point that needs to be addressed. Even if trusts and equitable assignments are effectively the same, could it be that the real illogic lies not in allowing indirect enforcement by the beneficiary of a trust of an otherwise unassignable obligation, but in the rule allowing anti-assignment clauses to invalidate equitable assignments in the first place? However non-physical or legally fictitious they may be, contractual rights are, after all, a kind of property. And since people can generally make what arrangements they like with their own property and alienate it as they think fit, it follows that we should recognise the rights of the beneficiary of a trust of a right, even if that right is purportedly unassignable.[58] Put another way, equity in recognising an assignment does not transfer the right from creditor to assignee, but rather leaves it in the assignee while acting *in personam* against the creditor to force the latter to give effect to the (property) rights he or she has purported to grant. And if this is right, then while an assignment in breach of an anti-assignment clause may be a breach of the creditor's contract with the debtor, there is no reason why it should affect the assignee's rights against the creditor.[59]

The answer to this, it is suggested, is threefold. First, although courts sometimes behave as if an equitable assignor indeed remains the owner of the assigned obligation,[60] with the assignee's rights lying against him or her alone,[61] it is suggested that equity's attitude to assignment went further than this and was more consistent with a view that the assignee should, as far as possible, be regarded as having the creditor's right actually transferred to him

[58] Lightman J expressed this thinking perfectly in *Don King*, above n 3, at 321: 'I can see no objection to a party to contracts involving skill and confidence or containing non-assignment provisions from becoming trustee of the benefit of being the contracting party as well as the benefit of the rights conferred.'

[59] A position well-expressed, though ultimately rejected, in RM Goode, 'Inalienable Rights?' (1979) 42 *MLR* 553. See also its discountenancing in *Linden Gardens*, above n 5. There are also signs of it in *Foamcrete Ltd v Thrust Engineering Ltd* [2002] BCC 221 (CA), on which see A Tettenborn, 'Prohibitions on Assignment–Again' [2001] *Lloyd's Maritime and Commercial Law Quarterly* 472.

[60] See, eg, *Warner Bros Records Inc v Rollgreen Ltd* [1976] QB 430 (CA), holding that the power to exercise an option under a contract remains exclusively in the original contractor despite equitable assignment. We ignore in this context the need to join the assignor in any action against the debtor, since this is today largely a formal requirement: see the authorities referred to above, n 30.

[61] Thus reproducing the classical view that equity does not subvert legal rights or titles, but rather recognises them and then puts constraints on their exercise: eg FW Maitland, *Equity: A Course of Lectures*, 2nd edn (Cambridge, Cambridge University Press, 1949) 106, 149.

or her. In particular, equity in pre-fusion days was prepared to act, not only against the assignor, but against the debtor directly, for instance by enjoining him or her from pleading set-offs arising after notice of assignment, or on occasion ordering him or her to account in equity (that is, to pay) to the assignee notwithstanding prior payment to the creditor.[62]

Second, even against the background of the rule that one can declare oneself trustee of any asset whatever, the present position of equity as regards assignments of non-assignable rights—that it will procure that, as between assignor and assignee, any benefits received by the assignor are handed to the assignee,[63] but that it will not ensure that the right is enforced against the obligor[64]—is not as unprincipled as it might seem. The reason is that there is no equity in granting a remedy that will have the effect of rendering nugatory a legitimate third party right,[65] in this case right of the debtor to decide with whom he or she is willing to enter into contractual relations. Since, as has been argued above, there is no difference worth the name between trusts and assignments, it seems to follow that exactly the same argument ought logically to apply to express trusts. Indeed, *Don King* and *Barbados Trust* themselves seem to go at least halfway towards this position. There are obiter dicta in both cases that suggest there may be at least some limits to the enforcement of an express trust of an obligation. In particular, both contemplate (but pointedly do not decide) that an express stipulation in a contract that no trust of it shall be created or recognised may be given effect against the purported beneficiary of any such trust.[66] If they are prepared to go that far, it remains to ask why they do not draw the logical conclusion and deny express trusts enforcement in so far as it would defeat the obvious expectations of the parties.

Third, and more generally, despite the underlying practice of English lawyers in classifying choses in action and their transfer as part of the law of property[67] (and hence, by transference, as an ordinary subject of the law

[62] *Malcolm v Scott* (1847) 6 Hare 570, 67 ER 1290 (ChD).

[63] *Glegg v Bromley* [1912] 3 KB 474 (CA).

[64] *Linden Gardens*, above n 5.

[65] For analogous situations, see, eg, equity's clear power to refuse specific performance where the effect of giving it would be to defeat a contractual stipulation entered into by the defendant in favour of a third party (*Warmington v Miller* [1973] QB 877 (CA)) and to refuse to lend its aid to an equitable chargee where, to the knowledge of the chargee, this would defeat a negative pledge clause previously agreed with a third party (*English & Scottish Mercantile Investment Co Ltd v Brunton* [1892] 2 QB 700 (CA)).

[66] See Lightman J in *Don King*, above n 3, at 321; Rix LJ in *Barbados Trust*, above n 8, at [88].

[67] As to this practice generally, and why it may have arisen, see Tolhurst, above n 1, at 15–17, 55–6, 62. In particular, it has made it easier for the law to accommodate a wide rule of prima facie transferability, with the attendant advantages to financing.

of trusts), there is also—as can be seen in the civil law approach[68]—a very distinct whiff of the law of obligations here. This is especially true where the right being transferred arises from a promise or other voluntary transaction on the part of somebody else.[69] In such cases, the concern is not simply with the relationship between a person and a thing: on the contrary, the other side of the right—the obligor's duties—are crucially in play. To whom, it is vital to know, does the obligor owe performance? With whom must he or she negotiate to cancel or reduce exposure? Can he or she choose to whom he or she is to be bound? And if so, then the terms of the obligation involved become, not some side-issue to the law of trusts or some matter as between debtor and creditor alone, but the central issue of legal policy.

It is therefore respectfully suggested that if, as seems to be the case, there is no difference of substance between equitable assignments and trusts of an obligation, one immediate conclusion is justified: in so far as assignments are prohibited by the nature of the obligation concerned, or its express terms, then there is no argument from logic in allowing that rule to be avoided by using the device of a trust. Nor is there much of a practical argument for it either. The justification for the English practice of allowing almost unlimited rein to anti-assignment clauses is thoroughly commercial: a debtor should be able to choose to make sure that if he or she goes to sleep next to Portia, he or she will not wake up entwined in the arms of Shylock,[70] whether by trust, assignment or anything else. The argument of the Court of Appeal in both *Don King* and *Barbados Trust* that (in effect) it was all the boxers' or Bank of Zambia's fault for not prohibiting both assignments and declarations of trust in so many words is on this count remarkably thin. Contracts and similar arrangements are, after all, to be interpreted in a reasonable, business-friendly and non-technical way, as the House of Lords has said[71] and Lightman J in *Don King* indeed accepts.[72] It is difficult, with respect, to think of a less businesslike, literal and nit-picking interpretation than one which attributes to contractors, as *Don King* and *Barbados Trust* do, an intention that a party to a contract should be protected from proceedings by a third party wearing the hat of an

[68] Tellingly, *cession* comes under the law of sale (effectively the law of obligations) in the Code Civil: see §§ 1689–1691. The relevant provisions in the Québec Civil Code (§§ 1637 ff) also appear under the law of obligations.

[69] I put it this way, rather than referring to the more natural 'contractual rights', because the same argument applies in respect of some obligations, such as those of professional competence, that can as a result of *Henderson v Merrett Syndicates Ltd* [1995] 2 AC 145 be classified as either contractual or tortious.

[70] See *Linden Gardens*, above n 5, and in Canada *Brio Beverages (BC) Inc v Koala Beverages Ltd*, above n 5.

[71] *Investors' Compensation Scheme Ltd v West Bromwich Building Society* [1998] 1 All ER 98 (HL) 114–15 (Lord Hoffmann).

[72] See *Don King*, above n 3, at 310–11.

assignee but not from proceedings by that same third party in the guise of a trust beneficiary. The House of Lords, or any other Commonwealth supreme court faced with this problem, as surely they will be, would do well to bear this in mind.

12

The Nature of Equitable Assignment and Anti-Assignment Clauses

C H THAM*

I. INTRODUCTION

ANTI-ASSIGNMENT CLAUSES are prevalent in modern commerce. This is due, perhaps, to the reluctance of contracting parties to deal with third parties with whom they may have had no prior relationship. The key English decision giving voice to this concern is that of the House of Lords in the appeals heard in *Linden Gardens Trust Ltd v Lenesta Sludge Disposals Ltd*[1] where it unanimously[2] took the view that the anti-assignment clause common to both appeals before it invalidated the purported equitable assignments[3] for value of the benefits of the building contracts in question. But in accepting that the assignments in the two appeals before them were ineffective, their Lordships opened up a 'black hole' of an incompensable loss,[4] since the would-be assignees were thereby merely third parties to the building contracts which the contractors

* Thanks are owed to Adrian Briggs, without whose encouragement this article would not have been written. All errors, however, remain mine alone.

[1] [1994] 1 AC 85 (HL) [*Linden Gardens*], being the conjoined appeals of *Linden Gardens Trust Ltd v Lenesta Sludge Disposals Ltd* (1990) 25 Con LR 28 (QB), rev'd (1992) 30 Con LR 1 (CA) and *St Martins Property Corporation Ltd v Sir Robert McAlpine Ltd* (1991) 25 Con LR 51 (QB), rev'd (1992) 30 Con LR 1 (CA) [*St Martins*].

[2] Lord Griffith's divergence was only on the issue of the manner in which the employer in one of the appeals (*St Martins*) could recover substantial damages from the building contractors.

[3] Though none of the speeches make this explicit, it is fairly clear that their Lordships were concerned with the effectiveness of the anti-assignment clauses *vis-à-vis* equitable assignments because in one case (*Linden Gardens*) notice of the assignments was never given and in the other case (*St Martins*) notice was only given long after the specific time under consideration. Since statutory assignment only occurs on receipt of written notice of the assignment, by implication both appeals must have been decided by reference to the equitable assignment of the relevant legal choses in action.

[4] It is probably true that this 'black hole' would still have manifested itself had there been no anti-assignment clause, or had effect not been given to the anti-assignment clauses: see *Darlington Borough Council v Wiltshier Northern Ltd* [1995] 1 WLR 68 (CA). But had

had breached. Having created the black hole, though, the majority in *Linden Gardens* promptly filled it back in by relying upon an extension of a common law exception developed by Lord Diplock in *The Albazero*.[5]

Leaving the science of black holes for another day, this article aims to re-examine the nature of assignments, in particular equitable assignments. It asks whether the general acceptance of anti-assignment clauses as effective to invalidate equitable assignments might not be due for an overhaul. The proposition in this article is that once we clarify what an equitable assignment of a contractual chose in action entails, the understanding of, inter alia, the proper role of an anti-assignment clause may well be ripe for change.[6]

II. EQUITABLE ASSIGNMENT OF BENEFITS

In England, that an anti-assignment clause may invalidate an otherwise effective assignment between promisee-assignor and third party-assignee is seemingly well accepted, following the decision of the House of Lords in *Linden Gardens*. There, the building contracts between the purported assignors and the obligors were on the Joint Contracts Tribunal ('JCT') standard form, and clause 17(1) prohibited assignments without the prior written consent of the obligor.[7] In Lord Browne-Wilkinson's view, 'clause 17(1) of the [JCT standard form] contract prohibited the assignment by the employer of the benefit of the contract. This, by itself, is fatal to the claim by [the assignee] in the *St Martins* case.'[8] Why? In Lord Browne-Wilkinson's view, (purported) assignments without the consent of the building contractors constituted a breach of clause 17,[9] and such assignments were ineffective to vest any causes of action in an assignee.[10]

Lord Browne-Wilkinson seems to have taken *Tom Shaw and Co v Moss Empires Ltd*[11] as standing for the proposition that a prohibition on assignment could invalidate the assignment as against the other party to the contract so as to prevent a transfer of the chose in action to a third

the House of Lords upheld the validity of the assignments in both appeals in *Linden Gardens*, the problem of the black hole would have become entirely academic.

[5] [1977] AC 774 (HL). Lord Griffith arrived at the same conclusion but through a different route.

[6] Assuming, of course, that the equitable assignment is not otherwise barred as being void for maintenance or champerty.

[7] The parties in *St Martins* were dealing with each other on the basis of the 1963 edn (July 1972 rev) form, whereas the parties in *Linden Gardens* were dealing with each other on the basis of the 1963 edn (July 1975 rev) form. Both versions of the JCT standard form had an identically worded clause 17(1).

[8] *Linden Gardens*, above n 1, at 103.

[9] *Ibid*, at 106.

[10] *Ibid*, at 107–108.

[11] (1908) 25 TLR 190 (KB).

party assignee.[12] The reasoning underlying this conclusion is, however, rather terse. There being neither statutory nor public policy reasons to render such prohibitory clauses void,[13] Lord Browne-Wilkinson read the existing authorities[14] as establishing that:

> an attempted assignment of contractual rights in breach of a contractual prohibition is ineffective to transfer such contractual rights. I regard the law as being satisfactorily settled in that sense. If the law were otherwise, it would defeat the legitimate commercial reason for inserting the contractual prohibition, viz., to ensure that the original parties to the contract are not brought into direct contractual relations with third parties.[15]

But does equitable assignment have such an effect—that is, to bring the original parties to the contract into direct contractual relations with third parties?[16] More specifically, does it bring the original obligor into direct contractual relations with a third party assignee? In *Warner Bros Records Inc v Rollgreen Ltd*, Sir John Pennycuick MR said not:

> Where there is a contract between A and B, and A makes an equitable but not a legal [or more accurately, a statutory] assignment of the benefit of that contract to C, this equitable assignment does not put C into a contractual relation with B, and, consequently, C is not in a position to exercise directly against B any right conferred by the contract on A. … [S]o long as the assignment remains equitable only, C has no more than a right in equity to require A to protect the interest which A has assigned[17]

Perhaps Lord Browne-Wilkinson was implicitly overruling the position taken in *Warner Bros*. But that would require one to assume that an attempted equitable assignment would have the effect of a 'transfer' of contractual rights. So three questions present themselves. First, does equitable assignment only operate to transfer contractual rights?[18] Second,

[12] *Linden Gardens*, above n 1, at 108.

[13] *Ibid*, at 106.

[14] Namely *Helstan Securities Ltd v Hertfordshire County Council* [1978] 3 All ER 262 (QB) in which Croom-Johnson J had held that a clause in a contract prohibiting the assignment of any benefits thereof was effective; *Reed Publishing Holdings Ltd v King's Reach Investments* (25 May 1983) CA transcript 121 which had dismissed an application to join an assignee as a party to the proceedings when the assignment was in breach of a prohibition against assignment; and *Re Turcan* (1888) 40 Ch D 5 (CA) which had proceeded on the basis of a valid declaration of trust on the basis that a contractual restriction on assignment was valid: see *Linden Gardens*, above n 1, at 106.

[15] *Linden Gardens*, above n 1, at 108.

[16] Lord Browne-Wilkinson mentions in passing that notice of the assignment was given to the obligors in one of the two appeals: *ibid*, at 101. But it is plain from the emphasis placed on the dates of the assignment in both cases, leaving aside any question of notice, that their Lordships' speeches were predicated on the assignments being equitable and not statutory: see n 3 above.

[17] [1976] QB 430 (CA) 445.

[18] This is principally examined in Part II.A, below.

does it transfer contractual rights at all?[19] Third, what does this tell us about the position taken by their Lordships in *Linden Gardens* on the effectiveness of anti-assignment clauses?[20]

A. Equitable Assignment Apart from Transfer

Given the commercial importance of assignment, it is a little surprising that it has not received much academic scrutiny. Within the last hundred years, only a handful of treatises have examined its inner workings.[21] Of particular note are Oshley Roy Marshall's *Assignment of Choses in Action*[22] and Joseph Starke's *Assignments of Choses in Action in Australia*,[23] and, much more recently, Marcus Smith's *Law of Assignment*[24] and Greg Tolhurst's *Assignment of Contractual Rights*.[25]

For Marshall and Tolhurst, the answer to the first question would be, 'no, an equitable assignment does not only operate as a transfer of contractual rights'. Both take the view that equitable assignment of a chose in action may operate as a contract between assignor and assignee,[26] as well as a conveyance[27] or transfer[28] of the chose in action. Now, if it were true that equitable assignment operated by way of transfer of contractual rights, the obligor's fear of being drawn involuntarily into direct contractual relations with a stranger might well be real. But even if the contract between assignor and assignee causes some form of 'transfer' of the

[19] This is principally examined in Part II.B, below.

[20] This is principally examined in Parts III.B and III.C, below.

[21] There have been, of course, numerous books which have dealt with the assignment of choses in action as a single chapter within the much broader topic of contract law. But by and large these treatments have not delved very far beneath the surface of judicial authority, perhaps unavoidably, given the constraints of dealing with such a complex topic within the confines of a more generalised discussion.

[22] OR Marshall, *The Assignment of Choses in Action* (London, Sir Isaac Pitman & Sons, 1950).

[23] JG Starke, *Assignments of Choses in Action in Australia* (Sydney, Butterworths, Sydney, 1972).

[24] M Smith, *The Law of Assignment: The Creation and Transfer of Choses in Action* (Oxford, Oxford University Press, 2007).

[25] G Tolhurst, *The Assignment of Contractual Rights* (Oxford, Hart Publishing, 2006).

[26] So long as the equitable assignment is by means of a contract of assignment, of course.

[27] Marshall, above n 22, at 119, in relation to what Marshall terms 'informal assignments' which are assignments operating '… neither by way of trust nor by way of contract …'.

[28] Tolhurst, above n 25, at [3.25]–[3.26], [4.07]–[4.08]. As to what Tolhurst means by a 'transfer' of a chose in action, see [3.10]. That conception incorporates notions of disposition, but is, ultimately, broad enough to encompass instances where there is no disposition of any interest in the chose at all: see text to and following n 54 below.

contractual rights in the chose in action assigned,[29] why should the impermissibility of the transfer affect the validity of the contract of assignment?[30]

Even accepting the conveyance or transfer analysis as accurate, neither Marshall nor Tolhurst suggests that equitable assignments may only take effect by means of a transfer. Marshall differentiates between three senses of assignment in equity,[31] namely an informal assignment,[32] assignments by way of contract,[33] and assignments by way of trust.[34] Tolhurst, in turn, re-christens assignments by way of contract as the 'remedial' model of assignment. Tracking Windeyer J's analysis in *Norman v Federal Commissioner of Taxation*,[35] he takes the view that this remedial model is 'not at odds with the idea that an assignment involves a transfer'.[36] So neither Marshall nor Tolhurst denies that an assignment may operate as a contract between assignor and assignee (nor could they, given the many cases which make this very point).[37] But as a contract between assignor and assignee, how might the obligor ever be brought into 'direct contractual relations with third parties', as Lord Browne-Wilkinson feared?

First, to allow this fear of the obligor to override the contractual bargain struck between assignor and assignee (where the assignment is for value)[38]

[29] This assumption is questioned in Part II.B, below.

[30] For the contract of assignment is surely the cause of the purported transfer, and to hold otherwise would be to allow the tail to wag the dog (assuming, of course, that the assignment was not intended to occur by way of gift).

[31] Marshall, above n 22, at 80–99.

[32] Typified by the events in *William Brandt's Sons & Co v Dunlop Rubber Co Ltd* [1905] AC 454 (HL). Smith suggests that such 'informal assignments' are, in fact, instances where a constructive trust has been imposed over the legal chose in action in question: Smith, above n 24, at [6.12].

[33] As noted by the Lord Chancellor in *Wright v Wright* (1750) 1 Ves Sen 409, 27 ER 1111 (Ch) 412.

[34] As noted by Cozens-Hardy LJ in *Fitzroy v Cave* [1905] 2 KB 364 (CA) 373. Marshall differentiates between the case where the assignor declares himself or herself to be a trustee (which would, in effect, be identical to the analysis undertaken above on the constitution of an express trust over the chose in action), the case where the assignor of the chose transfers it to trustees on trust for the beneficiary-assignee, and the case where the debtor is directed by the creditor to hold the chose in action in trust for the assignee.

[35] (1963) 109 CLR 9 (HCA).

[36] Tolhurst, above n 25, at [3.24], further expanded at [4.05]–[4.06].

[37] Lord Browne-Wilkinson expressly accepted this to be true: *Linden Gardens*, above n 1, at 108.

[38] The assignment under consideration in *Linden Gardens* was in the form of a deed, in consideration for £1 paid to the assignors on execution of the deed: *Linden Gardens* (CA), above n 1, at 9 (as recounted by Staughton LJ); *Linden Gardens* (HL), above n 1, at 100 (as recounted by Lord Browne-Wilkinson). In the other appeal, *St Martins*, the assignment of the assignor's contractual rights against the building contractors was also made by way of a deed. That deed, however, also encompassed the assignment of the assignor's proprietary interest in the development in question. Both Staughton LJ and Lord Browne-Wilkinson made it clear that consideration had been furnished in relation to the assignment of the proprietary interest in the development: *St Martins* (CA), above n 1, at 10; *Linden Gardens* (HL), above n 1, at 101. But even though there is no express mention of consideration in support of the

is little different from the ill-fated attempt of the Dunlop Pneumatic Tyre Company to impose a minimum price floor on third party re-sellers of its products, so forthrightly rejected by the House of Lords.[39] And certainly, it is nowhere plausibly suggested that the effect of an anti-assignment clause might be to prevent a disappointed contractual assignee from recovering damages from the assignor on account of losses arising from the failed contract of assignment.[40] To permit the anti-assignment clause to invalidate the contractual promise made by the assignor to the assignee would seem to be analogous, if not functionally equivalent, to the imposition of a contractual burden on a stranger to a contract. It is therefore difficult to see how we can avoid applying the general principle at common law that a contract may not impose burdens on anyone who is not party to it. Lord Browne-Wilkinson himself observed in *Linden Gardens* that:

> a prohibition on assignment normally only invalidates the assignment as against the other party to the contract so as to prevent a transfer of the chose in action: in the absence of the clearest words it cannot operate to invalidate the contract as between the assignor and the assignee and even then it may be ineffective on the grounds of public policy.[41]

This extract gives rise to the implication that even with an effective anti-assignment clause, the contract of assignment still remains effective as between assignor and assignee. All the anti-assignment clause then does, it seems, is to prevent a transfer of contractual rights owed to the assignor by the obligor to the assignee. And Lord Browne-Wilkinson makes this plain in the very next sentence, where he says:

> If on the other hand Darling J purported to hold [in *Tom Shaw v Moss Empires*[42]] that the contractual prohibition was *ineffective* to prevent [the assignor's] contractual rights against Moss Empire being transferred to Tom Shaw [the assignee], it is inconsistent with authority *and was wrongly decided*.[43]

Which is to say that Lord Browne-Wilkinson agreed that an anti-assignment clause only operates to invalidate the assignment as between the assignor (who is in breach of the anti-assignment clause) and the obligor (whose obligation it is that has been assigned), and presumably that it also prevents any possibility that the purported assignee might

assignment of the assignor's contractual rights against the building contractor, this was probably unnecessary since this formed part of the same deed, and in all likelihood, the consideration furnished by the assignees was intended to be in exchange for all the interests, proprietary or otherwise, which were to be furnished by the assignors under the deed.

[39] *Dunlop Pneumatic Tyre Co v Selfridge & Co Ltd* [1915] AC 847 (HL).

[40] Indeed, *Devefi Pty Ltd v Mateffy Pearl Nagy Pty Ltd* (1993) 113 ALR 225 (FCA) 236 tells us otherwise.

[41] Above n 1, at 108.

[42] Above n 11.

[43] *Linden Gardens*, above n 1, at 108 (emphasis added).

obtain via the assignment any direct contractual rights against the obligor which were previously owed to the assignor. However, this begs the question whether equitable assignment operates to 'transfer' contractual rights, owed by an obligor to the assignor, to the assignee. I suggest that it does not.

B. Equitable Assignment as Transfer: An Impossibility?

(i) The Supporting Role of Equity

Leaving aside statutory assignments pursuant to the Law of Property Act 1925,[44] the received wisdom is that common law made no general provision for the assignment of choses in action.[45] It just could not be done. Any 'transfer' of the chose would, at common law, have to be effected by novation (that is, a sort of surrender and regrant). It is also received wisdom that equity took a different view. In equity, it is said that choses in action are, in general, assignable.[46] If, however, 'transfers' of choses in action were not possible at law (as noted above), and if, 'equity follows the law,' how then may equitable assignment do what the law cannot, namely 'transfer' choses in action?

Lord Macnaghten said in *William Brandt's Sons & Co v Dunlop Rubber Co Ltd* that:

> [t]he [Supreme Court of Judicature Act 1873] does not forbid or destroy equitable assignments or impair their efficiency in the slightest degree.[47]

He went on to say that:

> [w]here the rules of equity and the rules of the common law conflict, the rules of equity are to prevail. … At law it was considered necessary that the debtor should enter into some engagement with the assignee [such as a novation]. That was never the rule in equity.[48]

There is little to quibble with in the first part of Lord Macnaghten's statement. But the latter part of his analysis requires more care. No specific authority is cited, although he may have had in mind the history of

[44] (UK) 15 & 16 Geo V c 20.

[45] Limited exceptions were made, eg, in relation to negotiable instruments and other 'documentary' choses in action.

[46] The editors of RP Meagher, *Meagher, Gummow & Lehane's Equity, Doctrines & Remedies*, 4th edn (Australia, Butterworths Lexis Nexis, 2002) 221 list the following exceptions to equity's benign attitude towards assignment: (i) bare rights of action; (ii) contracts involving personal skill and confidence; (iii) the salaries and pensions of certain public officers; and (iv) statutory rights which, expressly or impliedly, had been made un-assignable.

[47] Above n 32, at 461.

[48] *Ibid*.

jurisdictional disputes between the courts of common law and chancery that was ultimately resolved in favour of the chancery. And even though section 25 of the Supreme Court of Judicature Act 1873 provided for resolution of conflicts between rules of law and equity in relation to certain matters, in particular the assignment of choses in action (in section 25(6)),[49] Lord MacNaghten's observations ought surely to have been qualified by section 25(11):

> (11) Generally in all matters not herein-before particularly mentioned, in which there is any conflict or variance between the Rules of Equity and the Rules of the Common Law with reference to the same matter, the Rules of Equity shall prevail.

Now, even though section 25(6) only provides for a statutory means for assignment of choses in action, it would seem that the assignment of choses in action is, in consequence, a 'matter … hereinbefore mentioned' since it does address the question of assignment of choses in action. Such assignments are, therefore, matters which fall outside section 25(11): as particular mention had been made in section 25(6) as to the assignment of choses in action, there ought be no room for section 25(11) to operate so as to permit the rules of equity in relation to assignment of choses in action to 'prevail' over those at common law.

There is, indeed, nothing in the doctrine of equitable assignment for value which requires any notion of 'transfer' of the chose in action. Faced with a contractual agreement to assign a chose in action, equity, in its auxiliary jurisdiction, would grant specific performance against the assignor to perform the terms of the contract with the assignee to assign by compelling him or her to lend his or her name to any such proceeding that the assignee might wish to bring. There is no 'transfer'.[50] Rather, there is the creation of a new equitable interest.

Tolhurst explicitly acknowledges much of the above:

> [e]quity could not put in place rules for the transfer of legal rights. What it could and did do was take the position that, if a person intends to assign a legal right, that person should in certain circumstances … be bound by that act. Thus, equity acting *in personam* attaches to the conscience of the assignor and forces the assignor to lend his or her name to the suit at law.[51]

[49] (UK) 36 & 37 Vict c 66. Re-enacted as s 136(1) of the Law of Property Act 1925.

[50] As will be made clear, what the assignor does is not transfer any pre-existing rights. Rather, he or she creates a new right, assuming a new obligation recognised in equity owed to the assignee. See also Part II.B below.

[51] Above n 25, at [4.05].

Had the assignees in both appeals in *Linden Gardens* sought specific performance of the contract of assignment,[52] compelling the assignor to lend its name for the purposes of their bringing an action against the obligors to the choses in action assigned, no transfer of property would have been involved. At best, equity would recognise the 'assignee' as having some interest in the chose in action assigned, and that new interest is patently not the same as the chose in action assigned. Most obviously the new interest is not primarily directed at the obligor to the chose in action. The assignee's interest arises out of the 'assignment' and is directed at the 'assignor', against whom the order of specific performance is made. And the subsequent procedural changes which short-circuit the need to bring separate equitable proceedings against the assignor do not change the substantive logic underlying such assignments. So on this 'non-transfer' or 'contractual' view of equitable assignment, assignee and obligor are patently not brought into direct contractual relations with each other, and so the policy deployed by Lord Browne-Wilkinson to underpin his view in *Linden Gardens* ought not to have been of any concern.

For there to be an effective 'transfer' at common law, a novation is necessary.[53] And though the provisions for a statutory assignment left untouched the position in equity for assignments of a chose in action, such equitable assignment cannot 'prevail' over the common law requirement for a novation. So an equitable assignment cannot place the assignee in the same position as he or she would have been in had there been a novation. It may allow the equitable assignee to achieve much of what might have occurred had there been a novation, but complete congruency is not permissible, since to allow an equitable assignee to be in the position as if there had been a novation at law without having had to satisfy the common law's requirements would be to contradict the Judicature Act.

(ii) A Minimalist Usage of the Language of 'Transfer'

As set out above, an equitable assignment may operate as a contractual promise between assignor and assignee, in particular the promise by the assignor to lend his or her name to the assignee for the purposes of bringing legal proceedings against the obligor. The ubiquity of the verb 'to transfer' is, however, difficult to escape. So perhaps it is appropriate to continue using that verb, but only if we are careful about what is being 'transferred'.

[52] Presumably, this was not sought because the assignors in both appeals were only too happy to assist. But the point is that specific performance was available had it been sought.

[53] Or, following the enactment of the Judicature Act, compliance with the requirements to effect a statutory assignment pursuant to s 25(6) (or today s 136(1) of the Law of Property Act 1925).

I suggest in this article that an assignor 'transfers' a chose in action by handing over control of his or her freedom to decide whether or not to release the obligor from the contractual obligations under the chose in action assigned. This right to grant a release, of course, encompasses its corollary: the right to bring legal proceedings on the chose in action should there be a dispute as to the degree to which the obligor had completely performed the obligations under the chose.

The limited nature of equitable assignment is, in fact, explicitly recognised by Tolhurst:

> one might logically argue that although not all dispositions of rights involve transfers, a transfer requires a disposition. Such a restricted notion of title 'transfer' would not capture an equitable assignment of a legal right which at an analytical level merely creates and vests in the assignee an equitable interest. The assignor does not dispose of an equitable interest because when, one person holds 'the whole right of property,' no distinction is drawn between legal and equitable interests. Equitable interests are 'engrafted' or 'impressed' upon legal interests rather than 'carved out of' them.[54]

To fit this fact pattern, whereby an equitable interest in the chose in action is created by virtue of what takes place between the assignor and the assignee, Tolhurst stretches the meaning of 'transfer' to include this case.[55] But, with respect, if we wish to use the word 'transfer' in the context of an equitable assignment of a chose in action, we need to take care to recognise that it is a transfer that does not involve any disposition.

(iii) Is There Really Nothing More? Competing 'Maximalist' Views

Commentary on the question as to what is transferred by an equitable assignment of a legal chose appears to assume that the subject matter of the transfer is something more than just the right to bring legal proceedings. Smith observes that:

> [f]irst, and most narrowly, what is assigned could be no more than a right to sue in the name of the assignor. If this is what an equitable assignment of a legal chose is, it is remarkably similar in concept to the common law use of a power of attorney to enable an assignee of a chose to enforce the rights of the assignor.[56]

While acknowledging that there are cases[57] which hold that an equitable assignment merely gives the assignee the right to sue in the assignor's name,

[54] Above n 25, at [3.11] (citations omitted).
[55] *Ibid*.
[56] Smith, above n 24, at [6.07].
[57] Smith cites *Winch v Keeley* (1787) 1 TR 619, 99 ER 1284 (KB) 623; *Crouch v Credit Foncier of England* (1873) LR 8 QB 374, 380; *De Pothonier v De Mattos* (1858) El Bl & El 461, 120 ER 581 (KB) 467; *The Wasp* (1867) LR 1 A & E 367 (HC Adm) 368; *Walter & Sullivan Ltd v J Murphy & Sons Ltd* [1955] 2 QB 584 (CA) 588–9; *Three Rivers District*

Smith notes that they 'do not go so far as to state that this is the *only* right that the assignee acquires by virtue of the assignment'.[58] Smith posits that a second, alternative view is preferable, that is:

> the assignment could effect the transfer of the beneficial interest in the chose, leaving the assignor with nothing but the bare legal title. This implies a *substantive* effect on the assignor's rights in that the beneficial interest in the chose is separated from the legal title. In turn, this implies the creation of a trust.[59]

Smith argues that this second view is preferable because the first narrow conception of assignments would not be able to explain how, after notification of the assignment, the obligor may not be discharged from his contractual obligation unless he performs to the assignee.[60]

Tolhurst is of roughly the same opinion:

> Historically, for an equitable assignee of a legal right to enforce the legal right the action had to be brought by the assignee in the name of the assignor. If the assignor refused to allow its name to be used the assignee could, if the assignment was for valuable consideration, file a bill in equity and, upon giving an indemnity in costs, obtain an injunction allowing the assignor's name to be used in a suit at law to recover the assigned debt.…
>
> On the above analysis the equitable assignment of a legal right merely provided the assignee with a remedy against the assignor, that is, the assignment operated only between the assignor and assignee. What is obtained or 'assigned' from the assignor appears to be no more than a right to sue in the name of the assignor. … The effect of this contract analysis is that equity forces the assignor to perform its promise rather than leave the assignee with a remedy in damages. It is important to note what that promise is. Initially, the assignor had agreed immediately to assign a legal right to the assignee. Such an agreement was incapable of being performed at law[61] and therefore it cannot be the case that equity would uphold such contracts for assignment only if they were capable of being the subject of an order for specific performance. It appears that equity implied a promise that the assignor would lend its name to any suit against the obligor and held the assignor to this promise.[62]

Council v Bank of England [1996] QB 292 (CA) 299: see *ibid*, at 145 fn 16. Making much the same point, Tolhurst cites *Hammond v Messenger* (1838) 9 Sim 327, 59 ER 383 (Ch): above n 25, at 69 fn 19. He also cites *Long Leys Co Pty Ltd v Silkdale Pty Ltd* (1991) 5 Butterworths Property Reports 11,512, 11,518; *Showa Shoji Australia Pty Ltd v Oceanic Life Ltd* (1994) 34 NSWLR 548 (SC) 561; *Corin v Patton* (1990) 169 CLR 540 (HCA) 576; *Neave v Neave* [1926] Gaz LR 254, 256: above n 25, at 70 fn 20.

[58] Above n 24, at [6.07] (emphasis in original).

[59] *Ibid* (emphasis in original).

[60] *Ibid*, at [6.11]. This view might, however, require re-examination in light of the discussion of the cases relied upon in support: see Part II. C (iii), below.

[61] In the absence of any statutory assignment pursuant to the Judicature Act or the Law of Property Act 1925. Equitable assignment, of course, predates both statutes.

[62] Above n 25, at [4.05].

However, in his opinion, there were two problems with this analysis. First, there would be no need to resort to the language of assignment and vested rights.[63] Second, it would prevent the recognition of voluntary assignments.[64] Both problems would be obviated if, however, one accepted that equitable assignments were truly transfers, whereby the assignee would be 'vested with the ownership of rights existing between the assignor and the obligor by virtue of the contract between the assignor and obligor'.[65] And this state of affairs would arise, in Tolhurst's opinion, because:

> equity restated or progressed the remedial model so that the transfer was given effect to by reason of the remedies equity provided.
>
> Here, equity does not simply bind the conscience of the assignor and act *in personam*; rather, because the conscience of the assignor is bound equity fastens upon the subject property itself.…
>
> [A]n equitable interest is created and vested in the assignee, and that interest is the beneficial ownership of the legal right which is the subject of the assignment and the assignee is treated (in equity) as being owed the obligation. That is, in less precise terms (and in respect of contractual rights), the assignee obtains an interest in the contract that exists between the assignor and obligor. This is referred to here as the assignment analysis.[66]

Broadly speaking, therefore, both Smith and Tolhurst deny that an equitable assignment only takes effect by reference to the right to compel the assignor to lend its name for the assignee to bring legal proceedings against the obligor. Both suggest that in equity, some form of property interest is created and vested in the assignee by virtue of the equitable assignment, pointing out various problematic issues if such were not the case. A few observations may be made in reply.

First, there is no inherent inconsistency between the view that an equitable assignment involves only a 'transfer' of the right to grant a release (and, its corollary, the right to decide to bring legal proceedings on the chose in action) and the view that some equitable interest is created in the chose in action. That being the case, it is perfectly conceivable that the subject matter of the constructive trust[67] or the equitable interest[68] in the chose, as advocated by Smith and Tolhurst, merely relates to the right to

[63] *Ibid*, at [4.06].
[64] *Ibid*, at 70 fn 25. Presumably, this is because specific performance has no role in a case of a completed voluntary gift of a chose in action (there being no contract to specifically enforce). Nor would it have any role in relation to incomplete gifts (since equity does not lend its assistance to volunteers).
[65] *Ibid*, at [4.06].
[66] *Ibid*, at [4.07].
[67] On Smith's view.
[68] On Tolhurst's view.

grant a release and does not modify the terms specifying what is to constitute complete contractual performance by the obligor.

Second, though the focus of this article has been on equitable assignments for value, and on the specific enforceability of the contractual promise to assign, acceptance of this view does not preclude the possibility that some different basis may be relied upon to explain the phenomenon of voluntary equitable assignment, perhaps along the lines suggested by Tolhurst.[69]

Third, one ought not lose sight of contract principles pertaining to discharge by performance. Tolhurst distils the rules regulating *inter vivos* assignment of contractual rights as follows:

> the assignment of contractual rights is made up of a number of rules which overlap, make little sense as statements in their own right and appear to lack any general underlying and unifying principle. These rules are as follows:
>
> 1. An assignor can assign no greater right than it has nor can an assignee obtain a right greater than that held by the assignor.
>
> 2. Only non-personal contractual rights may be assigned.
>
> 3. It is not possible by assignment to increase or vary the obligations or burdens of an obligor.
>
> 4. It is possible to assign only rights and not obligations.
>
> 5. After receiving notice of an assignment, the obligor may not do anything to diminish the rights of the assignee.
>
> 6. An assignee can be in no better position than the assignor was prior to the assignment.
>
> 7. An obligor should be no worse off by virtue of an assignment.
>
> 8. An assignee takes subject to the equities.[70]

Tolhurst seeks to explain and rationalise these rules in Anglo-Australian law by proving two things. First, that all assignments, whether legal (statutory) or equitable do not just involve a transfer of rights; rather, for Tolhurst, assignment is transfer.[71] Second, that his Rules 5 to 8, 'to the extent that they focus on the obligor/assignee relationship rather than the assignor/assignee relationship, are explicable on this basis rather than the principle of transfer'.[72] In general, however, because Tolhurst begins with a conception of assignment as a form of transfer, the rule of *nemo dat* looms

[69] Alternatively, see CH Tham 'Careless Share Giving' (2006) 70 *Conveyancer and Property Lawyer* 411, 421ff.

[70] Above n 25, at [1.01].

[71] *Ibid*, at [1.02].

[72] *Ibid*.

large throughout his analysis. Thus, he takes the view that his Rules 1, 3, 6 and 8 are derived from the fundamental idea that a transferor can only transfer what he or she has at the time of transfer, and not anything more. This analysis is useful in certain contexts. However, for the purposes of this article, Tolhurst's starting points in relation to his treatment of Rule 3, as listed above, require closer scrutiny.

C. Equitable Assignment, (In)Variability of Contract Terms and Discharge by Performance

(i) Difficulties with the Maximalist Views

In relation to his Rule 3, Tolhurst makes the following assertion:

> [T]his rule is a clear adoption of the *nemo dat* rule. If the assignor cannot assign a right different from or greater than the one vested in him or her, then the effect of that assignment must be that the correlative obligation of the obligor is also not capable of variation by reason of the assignment.[73]

He then explains that:

> [it] may be noted that 'variation' here concerns variations of legal obligations rather than factual changes to performance. So long as the obligor is being asked to perform for the assignee the exact same legal obligation as that promised to the assignor, then the obligor cannot complain that there has been a variation to its obligation. It is accepted that there may be some increased inconvenience in fact by reason of an assignment.[74]

Both of these statements may be disputed. It is not obvious that the rule as to the non-variability of an obligor's contractual obligations derives from the *nemo dat* rule, for that non-variability plainly exists even where there is no question of assignment. Unless a contract expressly or impliedly permits the parties to a contract to modify the obligations undertaken by one party to the other, those obligations must be performed strictly in conformity with the agreed terms in order to discharge them by performance. Otherwise, the obligations as set out in such terms would be breached. This is really the result of the contractual principle that, unless otherwise provided for, '[t]he general rule is that a party to a contract must perform exactly what he undertook to do'.[75] It is not that assignment is incapable of effecting a variation because of *nemo dat*; rather, unless otherwise provided for, contractual provisions are inherently invariable.

[73] *Ibid*, at [1.03].
[74] *Ibid*, at 6.
[75] HG Beale (ed), *Chitty on Contracts*, 29th edn (London, Sweet & Maxwell, 2004) vol 1 [21–001].

Of greater concern is Tolhurst's second statement, requiring us to distinguish between 'variations of legal obligations' and 'factual changes to performance'. The former, he agrees, is impermissible and may not occur even where there has been a valid assignment. He reasons, however, that the latter is an inevitable and judicially sanctioned outcome. But how do we distinguish one from the other? To permit a 'factual change to performance' is to permit a 'variation of legal obligation.' They are one and the same because the question whether a factual performance has satisfied the requirements of the contract is determined by comparing what has in fact been done with what in law is required, the latter being determined through applying an appropriate construction of the relevant contractual terms. This is borne out by *Tolhurst v The Associated Portland Cement Manufacturers (1900) Ltd*.[76]

In that case, Alfred Tolhurst (the obligor) owned land at Northfleet in Kent containing extensive chalk quarries. He sold a small piece of this land to the Imperial Portland Cement Company Ltd (the obligee) which consolidated it with another piece of land so as to establish a factory to produce Portland cement. In January 1898, the obligor entered into a contract with the obligee to the effect that he would:

> for a term of fifty years, to be computed from the 25th day of December, 1897, or for such shorter period (not being less than thirty-five years) as he shall be possessed of chalk available and suitable for the manufacture of Portland cement, and capable of being quarried and got in the usual manner above water level, supply to the company, and the company will take and buy of the said [obligor] at least 750 tons per week, and so much more, if any, as the company shall require for the whole of their manufacture of Portland cement upon their said land.[77]

They contractually agreed that 1s 3d per ton would be paid in cash monthly for the chalk, and the average monthly payment for any year after 1898 was not to be less than £188.

As permitted by its constitutive documents, the obligee was voluntarily liquidated in 1900. Prior to that, it sold its entire undertaking, including the contract with the obligor, to the Associated Portland Cement Manufacturers (1900) Ltd (the assignee) via a deed of assignment.[78] The obligor took no part in the liquidation of the obligee, but appears to have supplied some chalk to the assignee, asking 2s per ton for it, on the basis that the 1898 contract had been brought to an end when he was given notice of the liquidation and sale of business. When the assignee declined to pay at the new rate, insisting instead that it was entitled to the benefit of the 1898

[76] [1903] AC 414 (HL) [*Portland Cement*].
[77] *Ibid*, at 418.
[78] This is made clear in the judgment of Collins MR in the court below: *Tolhurst v Associated Portland Cement Manufacturers (1900) Ltd* [1902] 2 KB 660 (CA) 665.

contract and to pay only 1s 3d per ton, the obligor brought an action against the assignee for the difference and for a declaration that he was no longer under any obligation to the obligee. The assignee and obligee cross-claimed for a declaration that the 1898 contract was still valid and binding, and that the assignee and the obligee were entitled to supply on the original terms.[79]

Mathew J at first instance found for the obligor on both the claim and the cross-claim. To his mind, it was:

> perfectly clear that [the assignee was] endeavouring to impose upon [the obligor] a contract into which he never entered, and that he [was] entitled to say that he did not make that contract. … Whenever, by reason of change of business or change of parties, the result of an assignment would be to impose upon one of the contracting parties a greater liability than he ever intended to assume, the contract [could not] be assigned.[80]

The Court of Appeal reversed Mathew J's decision, for the following reasons. First, it made no difference to the obligor whether payment of the price for chalk delivered pursuant to the 1898 contract was made personally by the obligee or some other party: that obligation was non-personal (in other words, payment of the price could be vicariously performed).[81] Second, the 1898 contract had not been repudiated, notwithstanding the obligee's liquidation and the sale of its business to the assignee.[82] Third, the 1898 contract only provided for a minimum level of supply—it did not provide for an upper limit of supply by reference to the expected scale of production. So there was no merit to the obligor's claim that he would, in effect, be put under a greater burden in light of the assignee's scale of business compared with that of the obligee.[83] Therefore, the obligor was not entitled to succeed on his claim for payment for chalk sold and delivered at a rate of 2s per ton, nor was he entitled to succeed on his application for a declaration that the 1898 contract was no longer binding upon him.

By a majority, the House of Lords upheld the decision of the Court of Appeal except on one critical point.[84] Delivering the leading judgment, Lord Macnaghten[85] took the view that although the 1898 contract only made express reference to the obligor and obligee, and made no express

[79] The particulars of the litigation and relief sought by the parties are culled from the headnote in the report of the first instance judgment of Mathew J: *Tolhurst v Associated Portland Cement Manufacturers (1900) Ltd* [1901] 2 KB 811, 813.

[80] *Ibid*, at 816–17.

[81] *Tolhurst v Associated Portland Cement Manufacturers (1900) Ltd*, above n 78, at 672, 679.

[82] *Ibid*, at 672–4, 678–9.

[83] *Ibid*, at 673, 680.

[84] Lord Robertson was the sole dissenting judge: *Portland Cement*, above n 76, at 421–2.

[85] With whom Lord Shand and the Earl of Halsbury LC agreed (albeit reluctantly).

provision for the assigns or successors in title of either, given that it was to last for at least 35 if not 50 years, a plain and literal reading of the contract would not do. Rather:

> when it is borne in mind that the [obligee] must have been induced to establish its works at Northfleet by the prospect of the advantages flowing from immediate connection with [the obligor's] quarries, and that the contract in substance amounts to a contract for the sale of all the chalk in those quarries by periodical deliveries … , it is plain that it could not have been within the contemplation of the parties that the company would lose the benefit of the contract if anything happened to [the obligor], or that [the obligor] would lose the benefit of the market which the contract provided for him at his very door in the event of [the obligee] parting with its undertaking, as it was authorized to do by its memorandum.[86]

Commercial reality required a more nuanced construction of the contract. Lord Macnaghten construed the 1898 contract as if it contained an interpretation clause saying that references to the obligor were to include his heirs, executors, administrators and assigns, owners and occupiers of the Northfleet quarries, and that references to the obligee were to include its successors and assigns, owners and occupiers of the cement works at Northfleet.[87] In consequence, the obligor was not entitled to recover payment for the chalk delivered to the assignee at the 2s rate. Nor was the obligee a necessary or proper party to the litigation.[88]

The key question was whether the obligor, Tolhurst, was bound to continue performing the obligation to deliver chalk, even though the obligee, the Imperial Portland Cement Co, was, practically speaking, no more.[89] In answer, Lord Macnaghten adopted a construction of the contract that expanded the meaning of the named obligor and obligee to include their successors and assigns.[90] It naturally followed that the obligor's principal objection to having to continue to perform his part of the contract, by reason of its impossibility of performance, had to fail.

[86] *Portland Cement*, above n 76, at 419. See also *Nokes v Doncaster Amalgamated Collieries* [1940] AC 1014, 1020, 1039.

[87] *Ibid*, at 420.

[88] *Ibid*, at 421. Lord Macnaghten did not explain why this was the case, but an explanation is hazarded in the text below at n 92.

[89] Lord Macnaghten remarked that though the joinder of the assignor would typically be expected, in his view, this was unnecessary where the assignor was, 'a mere name … without any executive or board of directors …': *ibid*, at 420–21.

[90] Tolhurst accepts that the obligor in *Portland Cement* 'had contracted to supply the needs of any person taking over the cement works who took an assignment of the right': above n 25, at [6.121]. But, taking the position that personal obligations are non-assignable, he seems to limit the significance of the construction point in *Portland Cement* to the question of whether the obligor was under a personal obligation to supply chalk to the assignors, the Imperial Portland Cement Co: above n 25, at [6.70].

(ii) Portland Cement *as a Case of Construction*

Portland Cement was not decided on assignment principles per se, but on a commercially sensible reading of the 1898 contract. Had the assignee been in existence on 12 December 1898, when the contract was formed, we would have been able to stop there, for on Lord Macnaghten's construction of the contract the contractual obligation of the obligor would then be owed not only to the obligee, but also to such person as might fall within that extended construction, such as the assignee. We would then be squarely within the realm of joint contractual promisees. But the facts of *Portland Cement* prevent us from applying that simple explanation. Even if the 1898 contract was read to include references to the obligee and its assigns or nominees, as at the formation of that contract on 12 December 1898, this particular assignee had not come into existence: it was only incorporated some time in 1900.[91] Therefore, it could not be said that, as at 12 December 1898, any contractual bargain had been constituted between the assignee and the obligor. Some other legal mechanism was required to fill the gap, but we do not need assignment to do the job.

The clue lies in Lord Macnaghten's explicit statement that the obligee was not a necessary party to the action.[92] A simple explanation could be found in the application of the doctrine of offer and acceptance, in particular, open offers. Approaching the 1898 contract from this perspective, and given the construction placed upon it by Lord Macnaghten, it is not difficult to take the view that the clause setting out the obligor's contractual responsibilities could be construed as an open offer to the obligee's assigns. The significance of the deed of assignment, naming the assignee, was that it enabled the assignee to fall within that category of persons to whom the obligor's offer to be bound was open. That offer was accepted when the assignee failed to reject that first delivery of chalk by the obligor after having been given notice of the assignment.

On this analysis, the sale of the business undertaking to the assignees did not 'transfer' the rights to performance of the 1898 contract: it only operated to allow the assignee to qualify as an offeree. Thus, *Portland Cement* is not a case about assignment as 'transfer' of contractual rights.[93] This should come as no surprise, for Lord Macnaghten's opening words in

[91] This is clearly stated to have been the case in the headnotes of the reports of *Portland Cement* at first instance, above n 79, at 812, and in the Court of Appeal, above n 78, at 662.

[92] *Portland Cement*, above n 76, at 421.

[93] This also explains, rather simply, why Lord Macnaghten took the view that the assignees would have been contractually obliged to order and purchase chalk from Tolhurst to the extent and at the rate stipulated in the 1898 contract, had the shoe been on the other foot: see *Portland Cement*, above n 76, at 420. Assignment theory cannot explain such assignment of burdens without resorting to rather complex theorising as to how, in some cases, burdens pass alongside benefits: see, eg, Tolhurst, above n 25, at [6.101]–[6.135], in particular, [6.116]–[6.121].

his speech in *Portland Cement* warned us as much: that though the case might be of great importance to the parties, from a legal point of view this case was of no importance at all.[94]

Lord Macnaghten's reading-in of the additional interpretation clause is central to his analysis, for had it not been read into the contract, it is arguable that the 'assignment' would have failed even if all statutory requirements had been satisfied.[95] As a statutory assignee, the assignee would have had transferred to it any remedy which the obligee might have been able to assert against the obligor once notice in writing of the assignment had been given. But without the interpretation applied by Lord Macnaghten, on a plain reading of the express terms of the contract, would the obligor have been in breach of the contract if he refused to deliver any more chalk? Of course not, since the named party to whom deliveries were to be made had been liquidated and, presumably therefore, it would no longer have been possible for the obligor to perform to the letter of the contract. This is the corollary of the principle of discharge by full performance.[96] Taking this principle seriously, if contractual performance requires performance to be rendered to a particular named person, delivery of goods or payment of that sum of money to another does not amount to precise performance and does not, therefore, discharge the contract.[97] Discharge, if it occurs, occurs because of some other doctrine.

(iii) Discharging a Debt by Making Payment to a Third Party

Admittedly, there is some authority for the proposition that in relation to the payment of debts, though the amount, time and place of payment are invariable, the party to whom the payment is to be made may be varied at the option of the creditor-obligee. Tolhurst provides the following example:

> if X owes A a debt of £100 payable at a certain place on a certain date, although A may be able to assign its right to the debt to a third party, that assignment alone cannot force X to pay the debt at another place or on another date.[98]

[94] Above n 76, at 416–17.

[95] At that point in time, the relevant provision would have been found in section 25 of the Judicature Act.

[96] If authority is needed, one might start by considering Sir Thomas Plumer MR's observation in *Goldsmid v Goldsmid* (1818) 1 Swans 211, 36 ER 361 (Ch) 219 (in relation to a specialty debt): 'Satisfaction supposes intention; it is something different from the subject of the contract, and substituted for it; … but with reference to performance, the question is, Has that identical act which the party contracted to do been done?' Allan Farnsworth put it this way: 'If a duty is fully performed, it is discharged. … The converse is equally clear. Nothing less than full performance operates as a discharge.': EA Farnsworth, *Farnsworth on Contracts* (Boston, Little, Brown and Co, 1990) vol 2 § 8.8.

[97] Unless we draw the distinction proposed by Tolhurst between 'legal' and 'factual' variations: see the quotation at text accompanying n 74 above.

[98] Above n 25, at [6.93].

For this, he principally relies[99] on the academic authority of *Corbin on Contracts*.[100] Both Tolhurst and Arthur Corbin leave open the possibility that the assignment may permit 'variations' in the identity of the party to whom X is to make payment in order to be discharged from his or her obligation of debt. Further, there is some old authority on which the editors of *Chitty on Contracts* rely in support of the proposition that if a creditor-obligee requests the debtor-obligor to pay the debt to a third party 'such a payment is equivalent to payment direct to the creditor, and is a good discharge of the debt'.[101] But those authorities, namely the old cases of *Roper v Bumford*[102] and *Page v Meek*,[103] require careful handling. *Roper v Bumford* is really a case of equitable set-off by agreement, and *Page v Meek* sets up no independent rule as to payment, but merely accepts as a point of pleading that payment to a third party at the direction of a creditor may be made as a special plea of payment. But from the face of the report, it would appear that the legal rationale for such pleading rested on the doctrine of accord and satisfaction. So it is significant that *Chitty on Contracts* merely states that the payment to a third party at the direction of a creditor is equivalent to payment direct to the creditor so as to discharge the debt. Though it is true that there may be a good discharge in such cases, the reason for such discharge does not lie in any equitable assignment per se, but by reference to other doctrines: set-off in one case, and accord and satisfaction in the other.

Similar treatment may be applied to some other cases that are often cited in support of the proposition that an obligor who makes payment or completes performance to the obligee-assignor instead of the assignee, following notice of an assignment, does so at his or her peril.[104] Yet the central case, *Brice v Bannister*,[105] may be read as being premised either on the lack of mutuality between the obligor and the assignee necessary to

[99] *Ibid*, at fn 754. Tolhurst also cites *RA Brierley Investments Ltd v Landmark Corp Ltd* (1966) 120 CLR 224, 231–2, 236; O Lando *et al*, *Principles of European Contract Law* (Dordrecht, Kluwer, 2003) Arts 11:306(1)–(2); Unidroit Principles of International Commercial Contracts (2004) Arts 9.1.8, 9.1.3. The relevance of these authorities, particularly the latter two, is somewhat limited. The relevance of *RA Brierley Investments Ltd v Landmark Corp Ltd* is also somewhat obscure as it deals with the rather distinct issue as to whether an entity that acquired shares in a company after a takeover offer had been made was entitled to accept such an offer.

[100] AL Corbin, *Corbin on Contracts* (St Paul, West Publishing, 1951) vol 4 § 868.

[101] Beale, above n 75, at [21–042].

[102] (1810) 3 Taunt 76, 128 ER 31 (CP).

[103] (1862) 3 B & S 258, 122 ER 98 (KB).

[104] See E Peel, *Treitel on the Law of Contract*, 12th edn (London, Sweet & Maxwell, 2007) 725–6, fn 95. This proposition was also doubted by Simon Brown LJ in *Deposit Protection Board v Barclays Bank Plc* [1994] 2 AC 373 (CA) 382, although the decision of the Court of Appeal was ultimately reversed by the House of Lords on other grounds: [1994] 2 AC 367 (HL).

[105] (1878) LR 3 QBD 569 (CA).

enable the obligor's advance payments to be statutorily set-off[106] against the debt arising from work done on a contract to build a boat, or on the basis of an insufficient degree of connection between such debts as to enable one to be set-off against the other in equity. As for *Jones v Farrell*,[107] it is not too difficult to read it as a case where there was an implicit finding that the contractual debt had been discharged not by performance to a third party at the behest of the creditor, but by accord and satisfaction, since the debtor in that case promised, in response to an order by the creditor to do so, that he would pay the third party debtor such sums as would become due under the agreement which was the subject of the assignment by the creditor. As for the remaining cases of *ex parte Nichols*[108] and *Durham Bros v Robertson*,[109] the former was only concerned with the issue of whether an equitable assignment of future debts executed prior to the bankruptcy of the assignee but after the occurrence of an act of bankruptcy was good against the assignee's trustee in bankruptcy. The latter was concerned with whether the assignment before the court was a statutory or an equitable one. Like many of the cases encountered in the course of this analysis, none of these cases stand as solid authority for the point for which it is commonly cited.

(iv) Equitable Assignment, the Invariability of Contract Terms and Discharge by Performance: a Synthesis

As this article has tried to demonstrate, the attempt to cast equitable assignment as a form of transfer creates unworkable inconsistencies if the concept of 'transfer' is applied indiscriminately. Cases such as *Roper v Bumford*,[110] *Page v Meek*,[111] *Brice v Bannister*,[112] or *Jones v Farrell*[113] do not compel one to accept that payment to the original creditor-obligee must be treated as being of no effect so far as discharge of the obligor's payment obligation is concerned. The final outcomes in those cases are explicable not by reference to some overriding principle peculiar to equitable assignment, but rather by reference to a variety of other legal

[106] Pursuant to s 13 of the Insolvent Debtors Relief Act 1729 (UK) 2 Geo II c 22 as amended by the Debtor's Relief Act 1735 (UK) 8 Geo II c 24 [the Statutes of Set-off]. In England, although the Statutes of Set-off were repealed in 1879 by s 2 of the Civil Procedure Acts Repeal Act, as the Court of Appeal has made clear, s 4(1)(b) of that statute preserved the rights of statutory set-off conferred by the Statutes of Set-off: see *Glencore Grain Ltd v Agros Trading Co* [1999] 2 Lloyd's Rep 410 (CA) 417.

[107] (1857) 1 De G & J 208, 44 ER 703 (Ch).

[108] (1883) 22 Ch D 782 (CA).

[109] [1898] 1 QB 765 (CA).

[110] Above n 102.

[111] Above n 103.

[112] Above n 105.

[113] Above n 107.

doctrines. The availability of means other than precise performance as defined in the contract between obligor and assignor to effect a discharge does not logically compel the conclusion that discharge by means of precise performance is no longer possible, and to insist on this may be excessively reductionist. This is entirely consistent with the analysis of *Portland Cement* set out above. For if equitable assignment varied the original obligation of payment so as to change the identity of the party to whom the contractual performance was owed, Lord Macnaghten's efforts to extend the meaning to be given to the term identifying the obligee in the contract would have been superfluous.

As noted above, it is axiomatic that a contractual obligation may only be discharged by precise and exact performance. The degree of exactitude will depend on the construction of the contractual obligation in question. However, where the party to whom delivery or payment is to be made is specifically named, unless the contract, as a matter of construction, permits such delivery or payment to be made to an entity other than that named party, it is arguable that such performance would be an unwarranted variation of the contract—the performance would be neither precise nor exact—and that obligation would be breached. This problem cannot be resolved by reference to assignment principles alone, and hence the need for Lord Macnaghten to have interpreted the contract in *Portland Cement* to include references to the obligee's assigns. In other contexts, the rules as to set-off, accord and satisfaction, or, perhaps, agency,[114] may have a role to play in appropriate circumstances so as to effect a discharge of the contract otherwise than by precise performance. The point is that equitable assignment does not address the question of whether a contractual obligation may be discharged by performance to the assignee. Such discharge, if it occurs, arises as a result of other legal doctrine. Equitable assignment, therefore, does rather less than is commonly claimed.

III. IMPLICATIONS OF THE MINIMALIST VIEW

A. Non-Assignability of Personal Obligations

To some extent, the points made above are not new. Writing in 1926 on the topic of assignment of contract rights, Corbin argued in favour of abandoning the language of 'alienability of choses in action':

> In continuing the discussion [on the alienability of choses in action] our first step should be to abandon altogether the term 'chose in action.' Its linguistic construction is faulty, in that its individual words lead one to think of something

[114] Alluded to by Chitty LJ in *Durham Bros v Robertson*, above n 109, at 770.

> very different from that which the expression as a whole now denotes. There is no 'chose' or *thing* or *res*. There is a right (or claim) against some person.[115]

This is, however, still a little too narrow. For if anything is 'transferred' or has been made the subject of an equitable interest that is vested in the 'assignee', it is the right to release the obligor from his or her contractual obligation.[116]

Realising what is 'transferred' in an equitable assignment of a contractual chose in action has certain implications. One of them requires us to re-examine Tolhurst's Rule 2, that 'only non-personal obligations can be assigned'. Given that an equitable assignment may operate by way of constitution of a trust, this rule is overstated. As Lightman J pointed out in *Don King Productions Inc v Warren*, there ought to be:

> no objection to a party to contracts involving skill and confidence [such as a personal obligation] or containing non-assignment provisions from becoming trustee of the benefit of being the contracting party as well as the benefit of the rights conferred.[117]

Agreeing with Corbin, the position in this article is that the same may be said where the equitable assignment operates otherwise than by constitution of a trust. Where it is only the right to grant a release that is 'transferred' by the equitable assignment, why should the fact that the legal proceedings relate to a 'personal' obligation invalidate such transfer? Corbin offers the following illustration:

> A contracts with B to act as B's valet. Surely, it will be said, B's right is so personal that it cannot be assigned. But no, the contrary is believed to be correct although no decision pro or con has been seen by the writer. By this statement it is not meant to say that the character of the service can in any way be changed by assignment. The right of B is that A shall act as B's valet, not that A shall act as valet for whom it may concern. Anyone ought to know that serving as valet to a cross, ill, miserly, old curmudgeon is not the same performance as serving a healthy, happy-go-lucky, generous, young prince. Therefore, when B assigns his right against A, he must assign it as it is. He cannot by assignment to C create in C a right that A shall act as C's valet. That would be a different right to a different performance. But B can assign to C the right that A shall serve as *B's* valet; and if A shall commit a breach it will be C who gets the damages measured by the value of the promised service.[118]

[115] AL Corbin, 'Assignment of Contract Rights' (1926) 74 *University of Pennsylvania Law Review* 207, 207.

[116] Even so, this is really an abuse of the language of transfer since the obligee may still bring legal proceedings against the obligor, such action being restrained only by the availability of injunctive relief on the application of the assignee.

[117] [2000] Ch 295, 321, aff'd [2000] Ch 291 (CA) [*Don King*].

[118] Above n 115, at 218 (emphasis in original).

Once the essence of what happens in an equitable assignment of a chose in action is made plain, the difficulty with the statement in Rule 2 becomes more obvious.[119] And if that is the case, how can an anti-assignment clause have any effect on this stripped-down and streamlined conception of equitable assignment? Surely it cannot.

B. Anti-Assignment Clauses

Where does this bring us in relation to Lord Browne-Wilkinson's application of the anti-assignment clause in *Linden Gardens*? With respect, it may be that a quite different view would have been taken of the efficacy of the anti-assignment clause in clause 17(1) had the limited nature of an equitable assignment been appreciated. As noted above, an equitable assignment may operate as a contract between assignor and assignee. That contract does not transfer (in the sense of a disposition) the benefit of the contractual performance due from the obligor, but 'transfers' the assignor's right to release the obligor from his or her contractual duties in consequence of the assignor's promise to lend his or her name to the assignee for the purposes of bringing an action on the subject matter of the assignment. As a corollary to the right to grant a release, the assignee would thereafter also have the right to cause a claim to be brought against the obligor should he or she default in the contractual performance.[120] And certainly, where there is only a contractual promise to assign, because 'equity looks on as done that which ought to be done' to give effect to the assignor's intentions, the courts of equity will recognise that an equitable interest in the thing assigned (which is the right to bring legal proceedings on the chose in action, and not the chose in action itself) has been created and

[119] Acknowledging Corbin's views, Tolhurst recognises that this is a possible mode of assignment: above n 25, at [6.75]. He describes this as being a case where 'if an assignment does not create privity of contract between the obligor and assignee, then arguably the effect of an assignment should be that the obligor continues to perform *to* the assignor but *for* the assignee; the performance is for the assignee as it is the assignee who owns the benefit of the right to performance and it is the assignee who can sue for breach of contract if the obligor fails to perform to the assignor. … [But t]he above possibility on its face takes no account of notice and dictates that this must be the effect of an assignment whether or not notice has been given. Clearly that is not the law. There is no doubt that upon receipt of notice the obligor can generally obtain a discharge only by accounting to the assignee and that prior to notice the obligor can obtain a discharge only by performing to the assignor' (emphasis in original). But, as has been explained elsewhere, this takes an overly reductionist view of how a contract might be discharged: see the discussion in Part II. C. (iii) and (iv), above.

[120] Tolhurst accepts that the term 'chose in action' may refer to 'a right enforceable by action or to the right of action itself': above n 25, at [2.02].

vested in the assignee.[121] As some others have described it, what equity does is to create rights to rights. It does not transfer the underlying right from one party to another.[122]

The scope of this equitable interest is necessarily very constrained if one is also to honour the maxim that 'equity follows the law'. The maxim that 'equity looks on as done that which ought to be done' cannot bring the assignee into any form of direct relationship with the obligor to the chose in action, as to do this would be to replicate in equity something which the common law would not have permitted in the absence of novation or statute. If an anti-assignment clause typically has no effect *vis-à-vis* an 'assignment' by way of trust,[123] that ought equally be true of an assignment in equity, since both, on this analysis, operate in a broadly similar fashion, although the specific constitutive intentions are distinct.[124]

Let us assume that the equitable assignment involves not a transfer (or at least not the same conception of transfer as one might apply to a statutory assignment or a novation) but the recognition of a new equitable interest in the chose in action, arising out of the specific enforceability of the assignor's promise to lend his or her name to the assignee for the purpose of bringing legal proceedings on the chose. If this is so, Lord Browne-Wilkinson's view that an anti-assignment clause may invalidate an equitable assignment cannot apply to an equitable assignment by means of a trust, since in the latter case, it is obvious that no direct contractual relationship is ever created between obligor and assignee. Yet because the processes by which equitable interests are created by the constitution of a trust or by means of an equitable assignment other than by constitution of a trust are so similar, if a typically worded anti-assignment clause is ineffective to prevent an assignment by way of trust (as was held to be the case in *Don King*),[125] the same should be true of an equitable assignment.

[121] Once executed consideration for the assignment has been furnished to the assignee. The clearest statement of this principle may be found in Meagher, above n 46, at [6–050]. But the same point is repeated in Tolhurst, above n 25, at 334–5. As to whether executed consideration had been furnished in the two appeals in *Linden Gardens* to the assignors in both appeals, see n 38 above.

[122] See R Chambers, *An Introduction to Property Law in Australia*, 2nd edn (Thomson Lawbook Co, Sydney, 2008) [13.90]; B McFarlane, *The Structure of Property Law* (Hart Publishing, Oxford, 2008) 70–74. Indeed, to underscore profound distinction between the common law and equitable conceptions of 'ownership,' McFarlane proposes that we abandon attempts to describe property rights as being 'equitable' and recognise that in equity, one has something rather different: a 'persistent right' (*ibid*, at 70–71).

[123] *Don King*, above n 117; approved by a majority of the Court of Appeal in *Barbados Trust Co Ltd v Bank of Zambia* [2007] 1 Lloyd's Rep 495 (CA) (Hooper LJ dissenting).

[124] Tony Oakley would characterise the equitable interest in the equitable assignee where the assignment was for value to have arisen out of a constructive trust: see AJ Oakley, *Constructive Trusts* 3rd edn (London, Sweet & Maxwell, 1997) ch 8. Smith follows suit: above n 24, at [6.12].

[125] Above n 117.

This, of course, inverts Andrew Tettenborn's persuasive arguments against the doctrinal soundness of *Don King*. Writing in response to the Court of Appeal's affirmation of Lightman J's decision, Tettenborn observed:

> Let it be granted that [the assignor's] rights under the management agreements were held on trust for the partnership and hence were partnership property. On that assumption, just what rights did the partnership have in respect of them? Had it the right to control their exercise, or to insist on their being used in a certain way? Could it have disposed of its beneficial interest to anyone it thought fit? If the answer to these questions is Yes, then there was an equitable assignment of the agreements in all but name.[126]

But that is precisely the point—in equity, one may effect an assignment both by means of a trust or otherwise. If the anti-assignment clause does not prevent an equitable assignment by constitution of a trust, it should likewise be unable to prevent an equitable assignment by means of the recognition of an equitable interest in the chose in action assigned. Although in the latter case equity is not recognising an equitable interest of a trustee-beneficiary, it nevertheless recognises a slightly different form of equitable interest, by reason of the specific enforceability of the assignor's contractual promise to lend its name to the assignee for the purposes of bringing legal proceedings on the chose in action assigned.

The Court of Appeal seems not to have shared Tettenborn's views.[127] I suggest that the force of his criticisms may be blunted if we accept that the concern with protecting the interests of an obligor may have been overstated. Tettenborn expands on the policy reasons for giving broader effect to anti-assignment clauses as follows:

> The object of allowing a contractor to stipulate that his obligations shall be unassignable is to allow him to obtain an assurance that he will have to deal with his co-contractor and no-one else, and that no-one other than that co-contractor will have the effective right to control the exercise of contractual rights against him.[128]

[126] A Tettenborn, 'Trusts and Unassignable Agreements–Again–*Don King Productions v Warren*' [1999] *Lloyd's Maritime and Commercial Law Quarterly* 352, 354. See also ch 11 of this book.

[127] The majority of the Court of Appeal expressly adopted Lightman J's analysis as to why anti-assignment clauses do not typically prevent parties from constituting themselves as trustees of the benefit in a chose in action in the recent case of *Barbados Trust Co Ltd v Bank of Zambia*, above n 123. The position of the third member of the Court of Appeal in this case, Hooper LJ, is less easy to discern since it seems he both approbated and reprobated Lightman J's analysis: *ibid*, at [139]. One therefore ought not to place too much weight on Hooper LJ's dissent. The point is discussed further in CH Tham, 'What Assignments of Causes of Action Are, and More—*Offer-Hoar v Larkstore*' [2007] *Lloyd's Maritime and Commercial Law Quarterly* 286, at text to fn 27.

[128] A Tettenborn, 'Trusts of Unassignable Agreements—*Don King Productions v Warren*' [1998] *Lloyd's Maritime and Commercial Law Quarterly* 498, 499. For a concurring view,

However, were there to have been no assignment, even in a case where the contractual promise is still wholly or partially executory, such assurance as the contractor might have had in the manner in which a co-contractor might exercise its contractual rights is entirely non-binding. This is particularly striking if we remember that an equitable assignment can effect no change in the legal obligations undertaken by the contractor to the co-contractor. Whether before or after assignment, the contractor's obligations under the chose in action remain exactly the same, unless some form of variation for consideration has been effected or some form of variability has been built into the contractual terms.[129] The question of whether the contractor is in breach of these obligations will still be measured against the terms of the original contract constituting the chose in action, and that is the precise question which the courts are to decide should legal proceedings be brought on the chose.[130]

As with the constitution of a trust, an equitable assignment involves no disposition, transfer or conveyance so as to bring the assignee and obligor into direct contractual relations with each other. Like the case in which a trust is constituted over a chose in action, an equitable assignment of a chose in action is an instance where equity is prepared to recognise the creation and vesting of an equitable interest in the chose in action in the assignee. In effect, a new equitable interest in the chose in action is created and vested in the assignee, rather than the chose in action being somehow transferred from one party to another. Or, in other words, a new equitable interest is 'engrafted' on to the chose in action. Therefore, even if clause 17(1) in *Linden Gardens* had had the effect of negating transfers of the benefit of the building contract without the consent of the obligor, it could have no effect on an equitable assignment of the benefit of the building contract operating via something more akin to what happens in the constitution of a trust of the chose.

C. Breaching an Anti-Assignment Clause: Ineffectiveness or Defeasibility?

The preceding discussion has attempted to demonstrate that an anti-assignment clause cannot prevent an assignor from effectively 'transferring', in equity, his or her right to grant a release of the contractual chose

see G McMeel, *The Construction of Contracts: Interpretation, Implication and Rectification* (Oxford, Oxford University Press, 2007) [16.77].

[129] Whether expressly, impliedly, or as a matter of construction. As argued above, the last is the true ratio of *Portland Cement*.

[130] Admittedly, there is one area where there is a discernible change as a result of an assignment, namely the promisor's expectations as to the likelihood that legal proceedings might be brought against him or her. However, as the discussion below in the text following n 137 shows, it is doubtful whether this is a legitimate interest deserving of judicial concern.

to an assignee of his or her choice. But we should not forget that an anti-assignment clause is, often, itself a contractual promise. The content of that promise is a matter of construction, but at the very least it is a promise by an obligee who is subject to an anti-assignment clause that he or she will not effect an assignment to a third party. So, although anti-assignment clauses do not, on the analysis above, prevent equitable assignments from occurring, the breach of the promise not to effect an assignment cannot be without consequence. Roy Goode identified at least one of these—where the breach of an anti-assignment clause constitutes a breach of contractual promise so severe that it gives the victim of the breach the right to elect to discharge the entire contract. This is confirmed by the law relating to the forfeiture of leases following a breach of a covenant against assignment.

As Goode recognised, one possible interpretation of an anti-assignment clause may be that it is a mere personal undertaking, 'the breach of which does not render the assignment ineffective against the debtor but merely exposes him to a claim for damages for breach of contract'.[131] In *Linden Gardens*, Lord Browne-Wilkinson accepted Goode's analysis that the legal effect of an anti-assignment clause was, to begin with, a matter of construction. He paraphrased Goode's itemisation[132] of the four possible constructions of an anti-assignment clause as follows:

> (1) that the term does not invalidate a purported assignment by A to C but gives rise only to a claim by B against A for damages for breach of the prohibition; (2) that the term precludes or invalidates any assignment by A to C (so as to entitle B to pay the debt to A) but not so as to preclude A from agreeing, as between himself and C, that he will account to C for what A receives from B … (3) that A is precluded not only from effectively assigning the contractual rights to C, but also from agreeing to account to C for the fruits of the contract when received by A from B; (4) that a purported assignment by A to C constitutes a repudiatory breach of condition entitling B not merely to refuse to pay C but also to refuse to pay A.[133]

Lord Browne-Wilkinson thought categories (1) and (4) to be unlikely,[134] and certainly he did not arrive at his conclusion on the effect of the anti-assignment clause before him on the basis of discharge by breach. But it is difficult to see just how clause 17(1) could ever have had the effect on

[131] RM Goode, 'Inalienable Rights?' (1979) 42 *MLR* 553, 554. This is also the general position adopted in the UNIDROIT Principles of International Contracts 2004: see, eg, the commentary to Art 9.1.3.

[132] *Ibid*.

[133] *Linden Gardens*, above n 1, at 104.

[134] *Ibid*.

the assignee's equitable interest in the chose in action assigned as Lord Browne-Wilkinson asserted it had, given the analysis put forward in Part II and Part III.B above.

It may be necessary, however, to go further. First, recognising that the source of an assignee's equitable interest in a legal chose in action 'transferred' to him or her by way of equitable assignment for value is the availability of specific performance, it should cause little surprise if a court should decide against making such an order in circumstances where an assignee had notice that the legal chose in action was subject to an anti-assignment clause. In such cases, the equities between obligor and assignee would hardly be equal.

Second, if the anti-assignment clause functions as a contractual promise in its own right, the breach of that promise may well bring about certain consequences which could discourage an obligor from breaching it, or a putative assignee from being involved in such a breach. Goode has already adverted to the possibility that such a breach could, in appropriate cases, entitle an obligor to discharge the contract. In cases where such discharge occurs early enough, preceding the date on which the obligor's contractual performance becomes due, this would be a reason for a putative assignee to think twice about his or her position. In such cases, even if the assignee had been wholly ignorant of the anti-assignment clause, there would be nothing left to specifically perform since the contractual obligations in the chose assigned would have been discharged by breach. What then of cases where discharge might conceivably occur after the contractual performance in the chose has fallen due? The simple discharge analysis will not help unless the obligor has been able to negotiate an anti-assignment clause which functions not merely as promise but also as a condition subsequent.[135] If so, the breach of the promise against assignment could well operate

[135] On this, Lord Steyn's explanation is helpful: 'Traditionally, a distinction is made between conditions precedent and conditions subsequent. Given that one is dealing with contingent as opposed to promissory conditions, one can for present purposes say that a fact is a condition precedent to a contract for the creation of which it is necessary; and that a fact is a condition subsequent to a contract that it extinguishes …': *Total Gas Marketing Ltd v Arco British Ltd* [1998] 2 Lloyd's Rep 209 (HL) 221. That said, although the breach of an anti-assignment clause may well be a fact that would extinguish the chose in action of which it is a part, that breach is also simultaneously the breach of a promise made by the assignor to the obligor in that chose. As such, it is not open to the assignor-in-breach to rely on the fact of its breach to assert that the chose in action has been entirely terminated: that option is left for the obligor to whom the promise of non-assignment is made. This is why the chose in action does not automatically determine on the occurrence of the condition subsequent. See also n 152 below.

retrospectively,[136] so as to deny legal effect to any contractual obligations that might have previously fallen due.[137]

The upshot, therefore, is that anti-assignment clauses may have the effect of deterring assignments in breach, but only as a matter of construction. No rule of law renders such assignments ineffective. If the policy reason behind such a rule of law is the desire to give effect to an obligor's wishes not to be brought into direct contractual relations with third parties such as assignees of the obligee, these policy grounds are weak. That policy presumably derives from the obligor's desire to minimise the risk of change in the manner and degree of supervision of the contractual performance and of the likelihood that legal proceedings would be brought against him or her. Yet even in a simple two-party case, there is little an obligor can do to prevent a hitherto quiescent and cooperative counter-party from undergoing the transformation from Dr Jekyll to Mr Hyde. Procedurally, the obligor may seek to convince the court that proceedings brought against him or her are unmeritorious. If so, such claims would be liable to be dismissed, presumably with an appropriate costs order. The obligor is expected to perform to the standards set in the contract, and the frivolity of any decision to litigate on that contract will be measured against that very standard. Given the degree to which the rule as to precise performance operates so as to render invariable the obligations undertaken within the contract, the fear that such contractually agreed standards and expectations may wildly fluctuate is misplaced. It is unclear how assignment could alter this.

It is true that the threat of legal proceedings brings about its own costs that may be irrecoverable. Apart from the reality that the courts are

[136] In *Total Gas Marketing Ltd v Arco British Ltd*, *ibid*, at 215 the House of Lords recognised that non-fulfilment of a condition subsequent could, as a matter of construction, have such an effect. Lord Slynn (with whom Lord Nolan, Lord Steyn and Lord Hope of Craighead concurred) said: 'If the provision in an agreement is of fundamental importance then the result either of a failure to perform it (it if is promissory) or of the event not happening or the act not being done (if it is a contingent condition or a condition precedent or a condition subsequent) *may* be that the contract either never comes into being or terminates. That may be so, whether the parties expressly say so or not. … To adapt the words of Mr Justice Maugham in *Re Sanderwell Park Colliery Co*, [1929] 1 Ch 277 at p 282 "the very existence of the mutual obligations is dependent on the performance of the condition." For completeness I would substitute "performance or fulfilment of the condition" for "performance of the condition"' (emphasis in original). Lord Hutton's speech makes much the same point, rejecting the submission that an unfulfilled contingent condition could only suspend the performance of obligations under the contract and not terminate it: *ibid*, at 226. Further, '… in regard to contingent conditions, it is not necessary for parties when stipulating for a condition precedent or a condition subsequent to spell out the consequences of non-occurrence of the condition: these are prima facie inherent in the use of such terms …', *ibid*, at 221 (Lord Steyn).

[137] This may be the proper basis underlying the operation of an 'anti-trusteeship' clause (as envisaged in *Don King*, above n 117) or an 'anti-equitable interest' clause, absent notice of such clauses.

reluctant to order indemnity costs on a routine basis, any threat of legal action causes the putative defendant to incur time, labour and effort in dealing with the claim. And if the proceedings were threatened by someone other than the original counter-party with whom some form of commercial understanding had been struck, existing calculations could be upset. But just as with the actual bringing of proceedings, the risk of such proceedings being brought and the calculations of those risks are, in the main, mere speculation. Unless there is a binding promise not to sue or to limit such a claim, or if there is an operative estoppel, there would be nothing to prevent the original obligee from upsetting all prior expectations. Why should the assignee be put in any worse position?[138] It is one thing for the law to impose burdens without an individual's consent (as in tort, criminal and revenue law, for example). It would be an entirely different thing for the law to allow individual private entities to make agreements to impose burdens on strangers to that agreement.

Through appropriate drafting, one might be able to create a situation where the court will refuse to grant specific performance to an assignee because it had notice of an anti-assignment clause. And even without such notice, an appropriately worded anti-assignment clause might well be construed as a condition subsequent, so as to release the obligor from all contractual obligations as yet unperformed, even if they are already due.[139] All of this is consistent with established learning in relation to assignments of leases in breach of covenant.

Millett LJ reminded us in *Hendry v Chartsearch Ltd* that:

> an assignment in breach of covenant is effective to vest the legal estate in the assignee: *Old Grovebury Manor Farm v Seymour Plant Sales & Hire (No 2)* … but the assignee takes a defeasible interest only which is liable to forfeiture for breach of covenant.[140]

A landlord's right to forfeit a lease may arise through (a) denial of the landlord's title, (b) breach of condition by the tenant, or (c) exercise of an express or implied forfeiture clause.[141] So even in relation to leases, there is no blanket rule that the breach of a covenant against assignment without obtaining the landlord's prior consent will render the assignment invalid: *Old Grovebury Manor Farm v Seymour Plant Sales & Hire (No 2)* tells us

[138] Certainly if there were such promises, limits or estoppels, on the analysis herein such restrictions would similarly bind the assignee, for an assignee only obtains the right to direct an obligee who has made an equitable assignment of the 'chose in action' in his or her contract with the obligor to bring proceedings against the obligor in relation to the obligor's contractual performance as owed to the assignor.

[139] Though this might well raise issues pertaining to the availability of equitable relief against forfeiture.

[140] [1998] CLC 1382 (CA) 1393–4, [1998] EWCA Civ 1276, [45].

[141] AJ Oakley, *Megarry's Manual of the Law of Real Property*, 8th edn (London, Sweet & Maxwell, 2002) 354–5.

otherwise.[142] Rather, the assignment is effective, though defeasible, and that defeasibility is not automatic. Furthermore, '[e]ven if the landlord has shown that he is treating the lease as forfeited, he may subsequently prevent himself from proceeding with the forfeiture if he waives the breach of covenant'.[143] The parallel with the model proposed in this article seems tolerably close, and there is no obvious reason why assignments of a chose in action in contravention of an anti-assignment clause ought to be treated any differently from assignments of leases.

This analogy was proposed but was rejected by the trial judge in *St Martins*.[144] Judge Bowsher QC took the view that this analogy was unhelpful:

> The law does not come to the assistance of assignees of leases who have taken in breach of covenant. On the contrary, the law assists the covenantee, even to the extent of holding that the covenant runs with the land and binds the assignees even though they are not mentioned in the covenant …
>
> Leases are in a different case to bare contractual promises. A lease contains rights in contract and rights of estate. The assignment of a lease passes an estate whereas an assignment of the benefit of a contract passes only a chose in action which is only of benefit if it can be enforced in the courts.[145]

In this, the trial judge was supported by Lord Browne-Wilkinson who observed:

> A lease is a hybrid, part contract, part property. So far as rights of alienation are concerned a lease has been treated as a species of property. Historically the law treated interests in land, both freehold and leasehold, as being capable of disposition and looked askance at any attempt to render them inalienable. … In contrast, the development of the law affecting the assignment of contractual rights was wholly different. It started from exactly the opposite position, viz., contractual rights were personal and not assignable. … It is therefore not surprising if the law applicable to assignment of contractual rights differs from that applicable to the assignment of leases.[146]

It is obvious that a lease is not entirely the same thing as a contractual chose in action. A lease is also a chattel real. But that difference is not relevant to our present discussion, and it is hard to see why it should lead to a legal distinction. In recognising that a lease is a chattel real, the law permits burdens to run with the estate in a lease so as to outflank the contractual rule against imposition of burdens on strangers to a contract. Although the doctrine of estates would also permit benefits to pass to those

[142] [1979] 1 WLR 1397 (CA).
[143] Oakley, above n 141, at 355.
[144] Above n 1.
[145] *Ibid*, at 62.
[146] *Linden Gardens*, above n 1, at 108–109.

in privity of estate, there is nothing in that to preclude such benefits from passing on the basis of contract. The rejection of the analogy with the position in relation to leases by both the trial judge in *St Martins* and Lord Browne-Wilkinson in *Linden Gardens* is therefore regrettable.

Millett LJ's judgment in *Hendry v Chartsearch Ltd* provides an alternative explanation. There, he said that:

> [i]n the case of a lease, the fact that an assignment in breach of covenant is effective to vest the term in the assignee means that it is too late to seek consent; the breach of covenant is complete and the lease is liable to forfeiture. That is not so in the case of the benefit of a contract. The assignment does not constitute a breach of contract and is without legal effect so far as the other party to the contract is concerned. It is not too late for the assignor to ask for consent. But the contract requires the assignor to obtain the prior consent of the other party; retrospective consent, if given, may operate as a waiver, but cannot amount to the consent require by the contract. The proper course if for the assignor to ask for consent to a new assignment and to wait until it is given or unreasonably refused to make it.[147]

There is little on the face of this statement, or in its context, to suggest that Millett LJ was only putting forward what he thought to be a plausible construction of the anti-assignment clause before him. The extract reads, perfectly naturally, as a generally applicable rationale that unless all the contractual pre-conditions to assignment of a chose in action have been satisfied, there is no assignment of the chose and no breach. Reiterating the point, Millett LJ said:

> The assignment which was made without the prior written consent of the defendants was effective as between assignor and assignee, but was ineffective as between the assignor and the defendants. The making of such an assignment did not put the assignor in breach of contract, let alone in repudiatory breach; it simply did not affect the [obligors'] legal position and could be disregarded by them with impunity.[148]

But that would seem to introduce a circularity[149]: if the breach of the contractual pre-condition as to non-assignability renders the assignment ineffective, then there is no breach; and if there is no breach, there ought to be no impediment to the effectiveness of the assignment; yet if that were

[147] Above n 140, at 1394.

[148] *Ibid*.

[149] This circularity has also been noted elsewhere: see GJ Tolhurst, 'The Efficacy of Contractual Provisions Prohibiting Assignment' (2004) 26 *Sydney Law Review* 161, at text to fn 86. In that article, Tolhurst goes on to argue that the presence of an anti-assignment clause, being a promise not to assign, '… rendered the contractual right personal and not capable of assignment': *ibid*, at text following fn 88. But, as the discussion above suggests, so long as we keep in mind the limited nature of the 'transfer' effected by an equitable assignment, there is no reason why personal obligations might not be assigned.

true, there would be a breach, and so on *ad infinitum*.[150] The only way out of this circularity is to recognise that an anti-assignment clause does not, per se, invalidate an assignment in breach of the promise not to assign, but breach of the promise not to assign may affect the chose in action which has been assigned. As suggested above,[151] one possible outcome arises where the anti-assignment clause is construed as a condition subsequent. If so, its breach will retrospectively nullify the contractual obligations contained in the chose, and that may be why the obligor, whose consent to the assignment has not been sought, is at liberty to ignore the assignment. But this does not invalidate the contract of assignment as between assignor and assignee. That assignment still stands, though as an assignment of subject matter which is defeasible at the option of the obligor.[152] Therefore, Millett LJ's assumption[153] that assignments of choses in action and leases operate differently in this respect may bear re-consideration.

Returning to *Linden Gardens*, it should be noted that there was no explicit consideration of these points, and therefore no finding that the breach of the promise not to assign had given rise to a right in the building contractors to discharge their obligations under the contract for breach of condition, much less a finding that such right of election had been exercised. Nor was there any finding as to the application of any express or implied contractual right to terminate the building contract. There was simply a breach of the anti-assignment clause. The proposition in this article is that non-compliance with an anti-assignment clause should have no impact on the validity of the contract of assignment as between the assignor and the assignee, the very point accepted to be the case in *Linden*

[150] One might try to finesse this, as Lord Browne-Wilkinson seems to have done, by suggesting that it is the purported assignment without the prior consent of the obligor in the *Linden Gardens* appeals that amounts to a breach, above n 1, at 106. But can one really 'purport' to do the impermissible? Is it really conceivable that contract law allows for the prescription of impossible intentions? Or might the rather less metaphysical model of defeasibility not be preferable? See also text to n 156 below.

[151] See text to and following n 135 above.

[152] Helpfully, in *Thompson v ASDA-MFI Group plc* [1988] Ch 241, 266 Scott J clarified the position set out in *New Zealand Shipping v Société des Ateliers et Chantiers de France* [1919] AC 1 (HL) and *Cheall v Association of Professional Executive Clerical and Computer Staff* [1983] 2 AC 180 (HL) [*Cheall*] as follows: 'In order to attract the principle that a party is not entitled to rely on his own acts as fulfilling a condition subsequent and bringing a contract to an end, the act must be a breach of duty and, per Lord Diplock [in *Cheall*]: "the duty must be one that is owed to the other party under that contract; breach of a duty whether contractual or non-contractual owed to a stranger to the contract does not suffice."' The promise not to assign contained in an anti-assignment clause is plainly owed to the obligor to the original contract and it amounts, therefore, to an obligation owed to the obligor. The obligor may therefore elect to ignore such a breach, if that suits his or her purposes. So on this analysis, it is not possible for the assignor-in-breach to unilaterally bring the chose in action to an end by virtue of assignment of the chose in breach of the anti-assignment promise therein.

[153] Above n 140, at 1394.

Gardens. The misstep appears to be the assumption that the breach of the anti-assignment clause would as a matter of law render ineffective assignments in breach thereof. Had the nature of equitable assignment, and the mechanism by which breaches of anti-assignment clauses affect the contractual obligations assigned, been better appreciated, it might have been easier for the House of Lords in *Linden Gardens* to conclude that breach of clause 17(1) should not have invalidated the assignment of the benefits of the building contracts in the two appeals. If so, there would have been no black hole to fill.

IV. CONCLUSION

Linden Gardens appears to have been argued and decided upon two assumptions: that equitable assignments have the effect of bringing a contractual obligor into direct contractual relations with third party assignees, and that the rules of assignment permit contracting parties to define limits as to when an assignment may be effective. But, as the preceding analysis has sought to demonstrate, the basis for those fears appears rather doubtful. That said, although this article takes the position that an anti-assignment clause does not render assignments in breach thereof ineffective as a matter of law, it recognises that it is possible that breach of an anti-assignment clause may, as a matter of construction, affect the chose in action assigned, and that this might give an obligor some avenues of redress should its unwillingness to be subject to the whims of a third party assignee be thwarted. However, that outcome is premised on defeasibility, as opposed to invalidity *ab initio*, of the assignment.

In opposition to this view of the limited effect of anti-assignment clauses, Tolhurst suggests that:

> the doctrinal efficacy of a prohibition on assignment lies in the ability of the parties to a contract to mould those rights they bring into existence and rob those rights of what would otherwise be their inherent transferability. Thus, although it is no doubt correct to suggest that contract operates through promises, it arguably does not follow that any prohibition on assignment, no matter how it is drafted, must amount to a mere promise not to assign. This would no doubt be the case if a prohibition operated only as a matter of contract, but it is not the case where the prohibition is intended to characterise the chose.[154]

With respect, this analysis downplays the significance of a breach of the contractual promise contained in an anti-assignment clause, focusing instead on the theory that an appropriately worded anti-assignment clause

[154] Tolhurst, above n 25, at [6.83].

can and does make assignment in breach thereof impossible, because it qualifies the 'transferability' of that which is sought to be assigned. If this view is applied to equitable assignment, it is questionable. As the analysis above has sought to show, at law, legal choses in action are inherently non-transferable if, by that, we mean the 'transfer' of an obligor's contractual obligations, such that it is to perform for the assignee instead of the assignor. In the absence of novation, this type of 'transfer' can only occur if the contract is construed so as to permit substitution or variation of the party to whom contractual performance is to be rendered. In some circumstances, such a variation may occur through the process of accord and satisfaction, and in some others the rules of set-off may permit us to conclude that a discharge of the original contractual obligation has been effected. But in none of these circumstances can it be said that the original obligation was discharged by performance.

Equitable assignment takes its place within this crowded space but does not operate in like manner. Rather, (a) if the assignor may be subject to an order of specific performance in relation to the exercise of his or her right to release the obligor from the contractual performance; then (b) the assignee may be said to have an equitable interest in the chose in action. That, loosely speaking, may well be described as a 'transfer of the right to grant a release'.[155]

There may be a view that all of the arguments above to refute 'ineffectiveness' of an assignment in favour of 'defeasibility' where the assignment has been made in breach of an anti-assignment clause like clause 17(1) is merely semantic—that we can safely short-circuit the process by saying that there are instances where breach of a clause will render an assignment ineffective. But in an echo of Millett LJ's rationalisation in *Hendry v Chartsearch Ltd*,[156] that runs the risk of circularity. Surely the law can do better. Quite apart from the attractions of enhancing the internal coherence of this corner of the law, there is advantage in reducing the legal risk posed by anti-assignment clauses to those who would be prepared to purchase such assets. For with a reduction in the legal risk, the discount applied to reflect that risk must, over time, reduce as well. Assignees face less risk, the assignors get more in exchange for what they are giving up, and obligors face no change whatsoever in the nature and extent of their legal obligations. What is there not to like about that picture?

[155] Together with its corollary, the right to bring legal proceedings (being the converse of a release).

[156] Above n 140.

13

Coming to Terms with The Great Peace *in Common Mistake*

KELVIN F K LOW*

I. INTRODUCTION

TO AN OUTSIDER, the predominantly adverse reaction of academic scholars to the decision of the English Court of Appeal in *Great Peace Shipping Ltd v Tsavliris Salvage* must seem very odd.[1] After all, the court sought to lay to rest the much-criticised decision of Denning LJ in *Solle v Butcher*.[2] It must surely be difficult to understand how a decision to overrule a mistaken decision can itself be mistaken. Perhaps the reaction to *The Great Peace* demonstrates a mellowing of academic opinion over Denning LJ's excesses in *Solle v Butcher*. Perhaps it is simply in the nature of academics to be contrarian, spurred on no doubt in recent years by an increasingly questionable pressure to publish more than can humanly be read. While *The Great Peace* cannot be regarded as providing complete and perfect justice, as the court itself readily admits,[3] it is questionable if the best way forward is to reverse *The Great Peace* rather than build upon it.

The Great Peace and *Solle v Butcher* are, of course, concerned with a particular species of mistake in contracting known as common mistake. Contractual mistake is of the 'common' variety when both parties to the

* Thanks are owed to Derek Davies and Francis Reynolds for their comments on a much earlier draft of this article when it took the form of an article entitled 'The Role of Equity in Mistake' that was presented at the 'Exploring Contract Law' symposium hosted by the University of Western Ontario in January 2008. I would also like to thank Mindy Chen-Wishart and Chee Ho Tham for helpfully pointing out that in its original form, that article was far too broad in scope and overly ambitious to boot, something which seems obvious in hindsight. This article focuses on the part of that earlier article which deals with common mistake in contract. My gratitude must also go out to Jason Neyers for the invitation to participate in a very lively and enlightening symposium as well as my colleague, Lusina Ho, for referring me to Jason.

[1] [2003] QB 679 (CA) [*The Great Peace*].

[2] [1950] 1 KB 671 (CA).

[3] *The Great Peace*, above n 1, at [161].

contract share the same mistake. Although the label is sometimes also used to refer to cases where both parties are contracting at cross-purposes,[4] this other variety of mistake is perhaps best referred to by its alias 'mutual mistake' to avoid confusion. To understand the controversy over the overruling of *Solle v Butcher* by *The Great Peace*, it is necessary to study two other cases, amongst others, in some depth—the decision of the House of Lords in *Bell v Lever Bros Ltd*[5] and the decision of Steyn J in *Associated Japanese Bank (International) Ltd v Crédit du Nord SA*.[6] These four cases are the most important landmarks in the development of the law of common mistake and in spite of doubts raised in *The Great Peace*, both *Solle v Butcher* and *Associated Japanese Bank* continue to feature prominently in contract textbooks.[7] However, rather than continue to look back to the past, I would suggest that *The Great Peace* provides us with the opportunity to look forward, a venture that I propose will prove more fruitful.

II. A ROUGH SKETCH OF THE LAW'S DEVELOPMENT

The acknowledged starting point for any study of the law of common mistake is the decision of the House of Lords in *Bell v Lever Bros Ltd*, itself a notoriously difficult case to understand. Very briefly, the rather odd facts of the case were as follows. The contract at the heart of the dispute is that which Lever Bros Ltd entered into with Ernest Hyslop Bell and Walter Edward Snelling to terminate Lever Bros' employment of the pair as chairman and vice-chairman respectively of Niger Company, Lever Bros' subsidiary. Lever Bros entered into this contract not knowing that they could have terminated the pair's employment contracts without payment since both had committed serious breaches of these contracts. To the surprise of Lever Bros, Bell and Snelling testified that they had forgotten

[4] As epitomised by *Raffles v Wichelhaus* (1864) 2 Hurl & C 906, 159 ER 375 (Ex Ct) wherein the parties purported to contract to buy and sell 125 bales of Indian cotton that would arrive in Liverpool on the ship *Peerless* from Bombay. It transpired that two ships named *Peerless* were to arrive in Liverpool from Bombay, one departing in October and the other departing in December. The buyer thought that the contract referred to the former whereas the seller thought that the contract referred to the latter. In light of these facts, the court concluded that a contract had not arisen.

[5] [1932] AC 161 (HL).

[6] [1989] 1 WLR 255 (QB) [*Associated Japanese Bank*].

[7] See, eg, HG Beale, *Chitty on Contracts, Volume I: General Principles*, 29th edn (London, Sweet & Maxwell, 2004) 5–026–5–047; E Peel, *Treitel: The Law of Contract*, 12th edn (London, Sweet & Maxwell, 2007) 328–30; SA Smith, *Atiyah's Introduction to the Law of Contract*, 6th edn (Oxford, Clarendon Press, 2005) 180–81. See also J Cartwright, *Misrepresentation, Mistake and Non-Disclosure*, 2nd edn (London, Sweet & Maxwell, 2007) 15.27–15.29.

about their breaches of duty when they entered into the termination agreements, setting the stage for an argument on common mistake.[8]

It is a perilous task to attempt to derive a ratio from *Bell v Lever Bros Ltd* but two points may be stated with some certainty. First, it is not every common mistake which will affect the validity of a contract but only extreme cases. Although various epithets can and have been used to describe the type of common mistake required to be established, it is really difficult to do better than 'fundamental'.[9] It is debateable what 'fundamental' means but the majority of the House of Lords in *Bell v Lever Bros Ltd* set an exceedingly high bar for passing the test given the unquestionable seriousness of the mistake in the case. The majority therefore appears to suggest that the scope of the nascent doctrine of common mistake is an exceedingly narrow one. Indeed, it paints so narrow a picture as to call upon itself criticism that it has been wrongly decided on its facts.[10] Secondly, the effect of common mistake at common law is to nullify consent and render the contract void *ab initio*. To understand this aspect of the decision, it is necessary to appreciate the background in which the case was decided. Early cases of mistake, it has been observed, tended not to reason from a doctrine of mistake, but either from an implication of a condition precedent as to the absence of the mistake[11] or on the basis of total failure of consideration.[12] 'This was an era when the parties' agreement was everything.'[13] Although the judges in *Bell v Lever Bros Ltd* do refer to common mistake *simpliciter*, traces of the earlier 'implied term' theory of common mistake can still be found in the case.[14] Under the influence of the 'implied term' theory of common mistake, the conclusion that any operative common mistake must render the contract void *ab initio* seems inevitable. If the doctrine of common mistake is to be alternatively analysed in terms of implied conditions precedent, then the result of the

[8] C MacMillan, 'How Temptation Led to Mistake: An Explanation of *Bell v Lever Bros Ltd*' (2003) 119 *LQR* 625, 645–6.

[9] JC Smith, 'Contracts—Mistake, Frustration and Implied Terms' (1994) 110 *LQR* 400, 401.

[10] *Ibid*, at 414–15.

[11] AWB Simpson, 'Innovation in Nineteenth Century Contract Law' (1975) 91 *LQR* 247, 268; *Barr v Gibson* (1838) 3 M & W 390, 150 ER 1196; *Couturier v Hastie* (1856) 5 HLC 673, 10 ER 1065; *Pritchard v Merchant's and Tradesman's Mutual Life-Assurance Society* (1858) 3 CB(NS) 622, 140 ER 885.

[12] *Strickland v Turner* (1852) 7 Exch 208, 155 ER 919.

[13] J Cartwright, 'The Rise and Fall of Mistake in the English Law of Contract' in R Sefton-Green (ed), *Mistake, Fraud and Duties to Inform in European Contract Law* (Cambridge, Cambridge University Press, 2005) 65, 73. Not only were common mistake and frustration attributed to the parties' presumed agreement, so too were the proper law of the contract (*Lloyd v Guibert* (1865) LR 1 QB 115, 120–21) and implied terms (*The Moorcock* (1889) 14 PD 64 (CA) 68).

[14] *Bell v Lever Bros Ltd*, above n 5, at 206 (Lord Warrington with whom Viscount Hailsham concurred) and at 224–7 (Lord Atkin with whom Lord Blanesburgh concurred).

doctrine must surely be to render the contract void *ab initio* since that is the basis upon which conditions precedent operate.

The narrowness of the common law doctrine of common mistake was seized upon by Denning LJ in *Solle v Butcher* to lay down a sister doctrine of common mistake in equity. In *Solle v Butcher*, the plaintiff, Godfrey Frank Solle, leased a flat from the defendant, Charles Butcher. The flat had previously been let at a rent of £140 per annum. Owing to damage by a mine in the war, the flat been repaired and substantially altered and both parties mistakenly thought that this freed the premises from rent control. They thus agreed to a rent of £250 per annum. After the relationship between the parties soured, Solle sought a declaration that the rent for the flat was £140 per annum and to recover his overpayments. Butcher responded by appealing to the court to set aside the lease for common mistake. The majority of the Court of Appeal expressed the view that the contract was voidable in equity for common mistake but that the contract ought to be rescinded upon terms being imposed in equity. This was because it transpired that Butcher could probably have charged close to £250 per annum for the flat if he only had complied with the statutory procedures which would have allowed him to make additions to the standard rent in respect of the improvements and structural alterations to the flat. *Solle v Butcher* marks the beginnings of the dissociation of the doctrine of common mistake from that of implied conditions precedent. According to Denning LJ, apart from cases where such a condition precedent can be implied into the contract, common mistake, however fundamental, did not render a contract void.[15] Therefore, the reality is that there is no separate and independent doctrine of common mistake at common law at all. Instead, common mistake as a separate doctrine, if it operated at all, operated purely in equity where the parties laboured under a common fundamental mistake, provided the party seeking to set aside the contract was not himself or herself at fault.[16] Where it operated, it rendered the contract voidable and not void and the courts retained a discretion to impose terms in awarding rescission.

Denning LJ's judgment in *Solle v Butcher* has been criticised for a number of different reasons. First, it has been suggested that Denning LJ's interpretation of the common law doctrine in *Bell v Lever Bros Ltd* was overly restrictive,[17] though it is arguable that this criticism is purely semantic since some of their Lordships in *Bell v Lever Bros Ltd* did appear to refer to the doctrine of common mistake interchangeably with that of

[15] Above n 2, at 691–3.
[16] *Ibid*, at 693.
[17] See *Associated Japanese Bank*, above n 6, at 267 where this suggestion is made.

implied conditions precedent.[18] Perhaps more importantly, it has been convincingly established that the cases from which Denning LJ derived the equitable jurisdiction in fact provided little authority for such a view.[19] His appeal to cases of misrepresentation and rectification for unilateral mistake do not support his view of the equitable jurisdiction in cases of common mistake since those situations clearly raise very different concerns.[20]

The apparent tension between *Bell v Lever Bros Ltd* and *Solle v Butcher* was addressed by Steyn J in *Associated Japanese Bank* in the third of our quartet of leading English cases. Here, Bennett contracted to sell and leaseback from the Associated Japanese Bank (International) Ltd four machines which did not exist. Crédit du Nord SA guaranteed Bennett's performance under the leaseback agreement. After making a single payment under the leaseback agreement, Bennett defaulted and the Associated Japanese Bank sued Crédit du Nord on the guarantee. The significance of *Associated Japanese Bank* is that it clearly dissociates the doctrine of common mistake from the implied condition precedent theory both at law and in equity. According to Steyn J:

> Logically, before one can turn to the rules as to mistake, whether at common law or in equity, one must first determine whether the contract itself, by express or implied condition precedent or otherwise, provides who bears the risk of the relevant mistake.[21] … Only if the contract is silent on the point, is there scope for invoking mistake … Where common law mistake has been pleaded, the court must first consider this plea. If the contract is held to be void, no question of mistake in equity arises. But, if the contract is held to be valid, a plea of mistake in equity may still have to be considered.[22]

Having thus differentiated for the first time common mistake at common law from the implied term theory, Steyn J accommodated *Solle v Butcher* by suggesting that the equitable jurisdiction to provide relief from common mistake was wider than that of the common law. This three stage test thus purports to reconcile *Solle v Butcher* with *Bell v Lever Bros Ltd*. Although Associated Japanese Bank's claim on the guarantee fell at the first hurdle, Steyn J suggested by way of dicta that even if that had not been the case, the contract of guarantee would in any event have been void at law for common mistake. In this respect, it rejects a narrow interpretation of *Bell v Lever Bros Ltd* which would restrict the common law doctrine to cases

[18] Above n 5, at 206 (Lord Warrington with whom Viscount Hailsham concurred) and at 224–7 (Lord Atkin with whom Lord Blanesburgh concurred).

[19] CJ Slade, 'The Myth of Mistake in the English Law of Contract' (1954) 70 *LQR* 385.

[20] J Cartwright, '*Solle v Butcher* and the Doctrine of Mistake in Contract' (1987) 103 *LQR* 594.

[21] Such risk allocation may also be the result of the rules of general law applicable to the contract: *William Sindall Plc v Cambridgeshire County Council* [1994] 1 WLR 1016 (CA) 1035 (Hoffmann LJ).

[22] *Associated Japanese Bank*, above n 6, at 268.

where there is a mistake as to the existence of the subject matter of the contract. On its facts, even though the machines did not exist, the machines were not themselves the subject matter of the contract of guarantee. Rather, the leaseback contract was the subject matter of the contract of guarantee and while the leaseback agreement was voidable for fraud, it was not void. So analysed, however, the similarities between *Associated Japanese Bank* and *Bell v Lever Bros Ltd* become obvious. 'In both, the subject-matter of the contract alleged to be affected by the mistake was an earlier contract; and in both that earlier contract was liable to be rescinded by one of the parties to the subsequent contract.'[23] Although there are differences between the two cases, they are not satisfactory grounds for distinguishing them.[24] Instead, *Bell v Lever Bros Ltd* is distinguished as 'a quite exceptional case' featuring peculiar facts.[25] Indeed, it would not be unfair to suggest that *Bell v Lever Bros Ltd* is far less influential than a decision of the House of Lords would ordinarily be and some may even suggest that Steyn J had merely paid lip service to the case.[26]

For almost 15 years, *Associated Japanese Bank* stood as the leading authority on the analytical approach to the subject of common mistake. Then the *Great Peace* sailed and an ill-advised piece of litigation changed the landscape of the subject. In *The Great Peace*, a ship, the *Cape Providence*, suffered serious structural damage in the Indian Ocean. When it learnt of this, the defendant, Tsavliris (International) Ltd, offered its salvage services, an offer which was accepted by the owners of the *Cape Providence*. As the tug allotted to the task was still a good five or six days away from the *Cape Providence*, Tsavliris approached a firm of London brokers who negotiated with the claimants, Great Peace Shipping Ltd, with a view to chartering the *Great Peace* to escort the *Cape Providence* until the tug arrived. A charter was concluded between the parties for a minimum of five days in the belief that the *Great Peace* was some 35 miles distant from, and the nearest vessel to, the *Cape Providence*. It transpired that the *Great Peace* was actually 410 miles away from the *Cape Providence* but Tsavliris did not immediately cancel the charter of the *Great Peace* upon learning of their mistake. Instead, it first sought to locate

[23] GH Treitel, 'Mistake in Contract' (1998) 104 *LQR* 501, 504.

[24] Treitel rejects, eg, the differences in the grounds of rescission and the fact that the earlier contract in *Bell v Lever Bros Ltd* involved the same parties whereas in *Associated Japanese Bank*, the earlier contract involved one party to the contract allegedly affected by mistake and a third party (*ibid*, at 504–505).

[25] Treitel, above n 23, at 505–507; See also MacMillan, above n 8.

[26] Treitel suggests that the difference in results in the two cases, whilst not exposing a direct conflict between them, demonstrates a conflict of policies (*ibid*, at 507). Whereas *Bell v Lever Bros Ltd* emphasised the need to respect the sanctity of contract, *Associated Japanese Bank* emphasised and gave effect to the reasonable expectations of honest men.

a nearer vessel. Fortuitously, another vessel chartered by the same charterers as the *Cape Providence*, the *Nordfarer*, happened to pass the *Cape Providence*, and Tsavliris entered into new arrangements for the *Nordfarer* to serve as escort for the *Cape Providence*. Tsavliris then sought to cancel the charter for the *Great Peace* some two hours after it had altered course to assist the *Cape Providence*. Whereas the agents of the *Great Peace* offered to persuade her owners to accept two days hire, Tsavliris refused to accept the conciliatory gesture. When they were sued upon the charter, Tsavliris argued that the entire contract was void for common mistake.

Considering their behaviour upon learning of the mistake, it should come as no surprise that their defence to the action failed, both at first instance before Toulson J[27] and on appeal to the Court of Appeal. Remarkably though, considering the clear facts of the case and the rules on precedent,[28] both Toulson J and the Court of Appeal sought to restate the English law of common mistake by overruling *Solle v Butcher* and consigning the equitable jurisdiction to intervene to the history books. In doing so, the Court of Appeal confirmed that the doctrine of common mistake at common law was not the result of some implied condition precedent but was a separate rule of law.[29] Whilst criticism on the basis of failure to adhere to the rules on precedent is not unexpected, it is somewhat odd, considering the pedigree of *Solle v Butcher*, to find that *The Great Peace* has received a generally frosty reception from both academics[30] and judges from other Commonwealth jurisdictions.[31] It appears that, despite initial hostility, *Solle v Butcher* has in recent years come to be regarded more favourably. Only two years prior to *The Great Peace*, Sir Christopher Staunton had expressed the view that *Solle v Butcher* 'can on occasion be the passport to a just result'.[32]

[27] *Great Peace Shipping Ltd v Tsavliris Salvage (International) Ltd* (2001) 151 NLJ 1696 (QBD) [*The Great Peace* (QB)].

[28] G McMeel, '"Equitable" Mistake Repudiated: The Demise of *Solle v Butcher*?' [2002] *Lloyd's Maritime and Commercial Law Quarterly* 449; SB Midwinter, 'The *Great Peace* and Precedent' (2003) 119 *LQR* 180; D Sheehan, 'Vitiation of Contracts for Mistake and Misrepresentation of Law' (2003) 11 *Restitution Law Review* 26, 33.

[29] *The Great Peace*, above n 1, at [73].

[30] FMB Reynolds, 'Reconsider the Contract Textbooks' (2003) 119 *LQR* 177; A Phang, 'Controversy in Common Mistake' [2003] *Conveyancer and Property Lawyer* 247; JD McCamus, 'Mistaken Assumptions in Equity: Sound Doctrine or Chimera?' (2004) 40 *Canadian Business Law Journal* 46.

[31] *Chwee Kin Keong v Digilandmall.com Pte Ltd* [2005] SGCA 2, [74]; *Miller Paving Ltd v B Gottardo Construction Ltd* (2007) 31 BLR (4th) 33 (Ont CA). On the former, see TM Yeo, '*Great Peace*: A Distant Disturbance' (2005) 121 *LQR* 393 and PW Lee, 'Unilateral Mistake in Law and Equity—*Solle v Butcher* Reinstated' (2006) 22 *Journal of Contract Law* 81.

[32] *West Sussex Properties Ltd v Chichester DC* [2000] NPC 74 (CA) [42] [*West Sussex Properties*].

III. CRITICISMS AND ANALYSIS

To the extent that the overruling of *Solle v Butcher* is justified by the suggestion that equitable intervention undermines the policy of the common law in maintaining commercial certainty,[33] the decision of *The Great Peace* cannot be sustained. It is 'elementary' that there must be a point in time at which a doctrine would have no precedent simply because it would be the first precedent to enunciate the principle concerned.[34] It is also not uncommon for equity to 'undermine' the common law by preferring different policies.[35] There is an admitted need in this area of the law to balance between the need to respect the sanctity of contract and the need to give effect to the reasonable expectations of honest men[36] and the suggestion that commercial certainty is of paramount concern cannot be seriously entertained. The Court of Appeal in *The Great Peace* did itself no favours by conceding that it would welcome greater flexibility than the common law permitted.[37] In suggesting that such development was beyond the common law and could only be provided through Parliamentary intervention, they failed to acknowledge that such statutory reform is hardly a priority for law reformers.[38] The Court of Appeal also suggests that, in resting 'fundamental' in equity upon notions of fairness,[39] the test in equity is far too uncertain,[40] but there is surely an element here of the common law pot calling the equitable kettle black.

The strongest case in favour of the reform undertaken by the Court of Appeal lies in the difficulty in disentangling the uncannily similar jurisdictions at common law and in equity, particularly since most cases will be addressed by the quite separate question of express or implied contractual allocations of risk. How does one distinguish between 'essentially different' at law from 'fundamental' in equity?[41] Neither test is completely precise, though imprecision in this context is neither fatal nor avoidable. Juxtaposed together, however, their interaction creates further, and unnecessary, uncertainty since both tests, occasionally identically worded, seek to capture the same idea. At some point when the parties' agreement runs out

[33] *The Great Peace*, above n 1, at [156].

[34] Phang, above n 30, at 251–2.

[35] For example, until the enactment of s 25(7) of the Judicature Act 1873 stipulations specifying the time of performance were generally regarded as 'of the essence' at common law, whereas the reverse was true in equity (*Parkin v Thorold* (1852) 16 Beav 59, 51 ER 698).

[36] Treitel, above n 23, at 507.

[37] *The Great Peace*, above n 1, at [161].

[38] Reynolds, above n 30, at 179.

[39] As Lord Denning MR envisaged in *Magee v Pennine Insurance Co Ltd* [1969] 2 QB 507 (CA) 514–15.

[40] *The Great Peace*, above n 1, at [138]. See also the criticism by Toulson J: 'Bluntly, the difficulty about this form of the doctrine is that it puts palm tree justice in place of party autonomy.' *The Great Peace* (QB), above n 27, at [120].

[41] *The Great Peace*, above n 1, at [131], [154].

(that is when the case cannot be decided on the basis of contractual allocation of risk, either express or implied), there may be occasion for the courts to relieve the parties of their bargain on the basis of common mistake. To balance between the need to protect the sanctity of contract and the reasonable expectations of honest men, such intervention will be rare and the mistake must be sufficiently serious before the courts will intervene. Whether the basis for intervention is described as 'essentially different' or 'fundamental', both epithets encapsulate the same idea—striking the appropriate balance. To suggest that the balance is differently struck at law and in equity, especially given the necessarily inexact terms of both tests, invites unnecessary uncertainty.

Despite the very real difficulty raised as to how a double test system with identical concerns and similar inexactitude can realistically be applied, critics of *The Great Peace* suggest that other concerns ought perhaps to take precedent. These number three. First, there is concern that the common law jurisdiction as expressed in *Bell v Lever Bros Ltd* and reformulated in *The Great Peace* is simply too narrow[42] and equity's intervention is perhaps necessary to supplement the law. Secondly, concerns have been raised about the potential for third party rights to be implicated by the effects of common mistake at law as the doctrine causes the contract to be avoided *ab initio* rather than be rendered merely voidable.[43] Finally, and perhaps most critically, the loss of remedial flexibility in equity is regarded as lamentable.[44]

A. The Limits of the Common Law

To suggest that the result in *Bell v Lever Bros Ltd* casts the doctrine of common mistake at common law in too narrow a light is not to make a particularly profound statement. It was convincingly demonstrated more than 10 years ago that the facts of the case could easily have been treated as a case of either *res sua* or *res extincta* (depending on whether one adopted the perspective of Lever Bros or that of Bell and Snelling) and on either view, the contract should have been void at law.[45] A more recent, and very telling, historical analysis of the case has cast further doubts as to the correctness of the case.[46] Although there are suggestions in *Bell v Lever Bros Ltd* that it was not clear, given Lever Bros anxiousness to terminate the two agreements as a result of the merger of the Niger with its rival, that Lever Bros would not have entered into the disputed contract to terminate

[42] McCamus, above n 30, at 77–8.
[43] Phang, above n 30, at 252–3; McCamus, *ibid*, at 79.
[44] Reynolds, above n 30, at 179; Phang, *ibid*, at 252–3; McCamus, above n 30, at 80–81.
[45] Smith, above n 9, at 414–15.
[46] MacMillan, above n 8.

the employment of Bell and Snelling even if they had known of the voidability of their employment contracts,[47] this seems altogether unreal when the facts are properly appreciated. It has been suggested, quite rightly, that by the time the case reached the House of Lords, the case had became detached from its facts.[48] Bell and Snelling's egregious breaches had transformed into foolish mistakes which caused no harm to Lever Bros. Instead of being recognised as managers of a huge corporate concern, Bell and Snelling are cast as mere 'servants'.[49] The focus turned from their breaches to their exceptional service to the Niger Company. The suit instituted by Lever Bros came to be viewed, not as an action concerned with commercial honesty,[50] but as an action by a greedy corporation bent on trying to save money.[51] It is also significant that the case was decided a decade before directors' duties came to be authoritatively expounded in *Regal (Hastings) Ltd v Gulliver*.[52] Viewed through more modern lenses, many of these concerns seem out of place. It is orthodox fiduciary law that breaches are not excused even if they resulted in no harm to the beneficiary. Indeed, they are not excused even if the breaches *themselves* resulted in the beneficiary making a tidy profit.[53] How then can it be relevant that the fiduciary had, apart from the breach, performed exemplary service? The suggestion that Bell and Snelling were mere 'servants' would today be regarded as so absurd that any lawyer making such a suggestion would likely be laughed out of court. As a result, *Bell v Lever Bros Ltd* is difficult to reconcile with more recently decided cases on the subject of fiduciary law such as *Item Software (UK) Ltd v Fassihi*.[54] *Bell v Lever Bros Ltd* is very much a decision coloured by concerns which are today irrelevant. However, most modern cases do little more than pay lip

[47] *Bell v Lever Bros Ltd*, above n 5, at 236 (Lord Thankerton).

[48] MacMillan, above n 8, at 650.

[49] See the various divergent examples constructed by Lord Atkin in his judgment to support his conclusion: *Bell v Lever Bros Ltd*, above n 5, at 228.

[50] As it had been at trial: see *Bell v Lever Bros Ltd*, above n 5, at 625 in which the closing submission of Lever Bros Ltd's counsel, Stuart Bevan KC, at trial is quoted: 'Do you think this action is brought for £30,000 or £20,000? … This action … was essential to bring, because if this sort of thing goes on and is permitted to go on, the whole fabric of commercial honesty and the whole structure of business becomes rotten.'

[51] *Bell v Lever Bros Ltd*, above n 5, at 213 (Lord Atkin). This suggestion does not accord with the real motivation for prosecuting the action. According to MacMillan, above n 8, at 640: 'It was unlikely to have been brought to recover the severance sums for the sake of the money alone. The existing Lever Brothers legal files disclose no attempts to settle the action. Lever Brothers budgeted £20,000 to prosecute the action. They persisted when the costs had risen and they estimated that the cost of the action would be £30,000 of which they might recover £15,000. In the end, they spent in excess of £40,000. This was a suit brought upon principle and not for profit.'

[52] [1967] 2 AC 134 (HL).

[53] *Boardman v Phipps* [1967] 2 AC 46 (HL).

[54] [2004] EWCA Civ 1244. Note, however, the criticisms of the case by A Berg, 'Fiduciary Duties: A Director's Duty to Disclose his Own Misconduct' (2005) 121 *LQR* 213.

service to *Bell v Lever Bros Ltd* in any event.[55] It may be best for *Bell v Lever Bros Ltd* to be laid to rest but the likelihood of litigation over common mistake recurring, much less reaching the House of Lords, seems remote at best. Lip service seems to be the tidiest solution in the interim. Certainly the recasting of the test for common mistake by the Court of Appeal in *The Great Peace* in terms equivalent to those applicable to both frustration and fundamental breach suggest that *Bell v Lever Bros Ltd* will cast nothing more than a pale shadow in the future.

However, this linking of the test for common mistake to that of frustration and fundamental breach has been criticised as a regression. It is suggested that it 'more or less returns the analysis to one of being required to determine whether the subject matter of the contract is no longer in existence, in some sense, and therefore cannot be delivered'.[56] Such a narrow test, it is said, will drive the courts to manipulate the concept of 'the subject matter of the agreement' so as to produce the desired result.[57] It is not self-evident that the Court of Appeal actually does this. It has never been clear that the oft-cited test at common law of 'essentially different' was meant to serve as anything more than a metaphor for a test that cannot be precisely stated and is intended to illustrate the need to balance between competing policy interests. In citing Diplock LJ's pioneering judgment in *Hongkong Fir Shipping Co Ltd v Kawasaki Kisen Kaisha Ltd* in which his Lordship opines that '[t]he test … has been stated in a number of metaphors all of which … amount to the same thing: does the occurrence of the event deprive [a party] of substantially the whole benefit which it was the intention of the parties as expressed in the contract that he should obtain as the consideration for [performance]?'[58] it is not evident that the Court of Appeal was returning common mistake to the Jurassic age where it was believed that only *res sua* or *res extincta* would render the contract void. If anything, this reformulation helpfully frees the law of common mistake from the vice-like grip of *Bell v Lever Bros Ltd*. The suggestion that the link to frustration and therefore impossibility calls for such an interpretation of *The Great Peace* takes the idea of impossibility in frustration far too literally. 'Impossible', it has been said, is 'something of a relative term'.[59] It cannot mean literal impossibility, else *Taylor v Caldwell*[60] should have been decided differently.[61] However, the English courts have generally resisted the temptation to transition to the more lenient

[55] See, eg, *Associated Japanese Bank*, above n 6.
[56] McCamus, above n 30, at 78.
[57] *Ibid*.
[58] [1962] 2 QB 26 (CA) 66.
[59] Peel, above n 7, at 940.
[60] (1863) 3 B & S 826, 122 ER 309.
[61] Peel, above n 7, at 940, citing LL Fuller and MA Eisenberg, *Basic Contract Law*, 3rd edn (St Paul, West Publishing, 1972) 801.

'impracticable' because, the test being vague enough as it is, it is necessary to drive home the message that sanctity of contract is a very real concern to the courts in balancing the competing policy interests. It is certainly questionable whether an open-textured rule drawn in terms of 'fundamental' or 'basic'[62] is either excluded by the terms of *The Great Peace* or would provide greater guidance.

The suggestion that American law is superior in this respect is questionable.[63] According to § 152 of the *Restatement (Second) of the Law of Contracts*, the test for common mistake is as follows:

> (1) Where a mistake of both parties at the time a contract was made as to a basic assumption on which the contract was made has a material effect on the agreed exchange of performances, the contract is voidable by the adversely affected party unless he bears the risk of the mistake under the rule stated in § 154.[64]

An English or Canadian lawyer would be alarmed by the label 'material' but it is stressed that 'material' in American means something quite different than 'material' in English.[65] This begs the question why it is superior? Both terms—'material' and 'impossible'—are relative and equally capable of manipulation. Both terms require further explanation as to the competing policies they seek to balance. In American law, the Commentary to § 152 provides:

> The mere fact that both parties are mistaken … does not, of itself, afford a reason for avoidance of the contract by the adversely affected party. Relief is only appropriate in situations where a mistake of both parties has such a material affect on the agreed exchange of performances as to upset the very basis for the contract … It is not enough for [a party] to prove that he would not have made the contract had it not been for the mistake. He must show that the resulting imbalance in the agreed exchange is so severe that he can not fairly be required to carry it out.[66]

It is not immediately evident that the American test or description encapsulates a markedly different idea or captures the same idea more effectively than Diplock LJ's test in *Hongkong Fir*. Surely then the choice of labels—between 'material' and 'impossible'—merely serves to highlight the message a particular legal system wishes to send to potential litigants as to which way the system leans in the application of a test as vague and open-textured as that which applies in the context of common mistake, frustration and fundamental breach. Are the courts suggesting that they lean in favour of sanctity of contract, thereby discouraging litigation; or

[62] McCamus, above n 30, at 78.
[63] *Ibid*, at 63–66, 76–79.
[64] *Restatement (Second) of the Law of Contracts* (1981) § 152.
[65] McCamus, above n 30, at 63–4.
[66] *Restatement (Second) of the Law of Contracts*, above n 64, at 386.

that they lean in favour of providing relief from the potentially very serious consequences of mistake, thereby encouraging litigation?[67] It is not self-evident that either view is preferable. Nor, given the rarity of cases of common mistake, does it seem practical for the courts to agonise over so slight a difference or academics to spill so much ink over it. Everyone will have his or her own preference as to a choice between the two labels but there are surely more important issues to mull over than the label to the already flexible test for common mistake. After all, the vast majority of cases will be decided on the basis of express or implied contractual allocations of risk.[68]

B. The Protection of Third Parties

The protection of innocent third parties is a concern which, at first blush, appears worthy of concern, but on careful examination proves to be much less deserving. The idea that innocent third parties need protecting was first pursued by Denning LJ in *Solle v Butcher*.[69] It is said that if the contract is void at common law, no property will pass to the original purchaser. *Nemo dat quod non habet*, the original purchaser will thus fail to pass title on to the innocent third party purchaser. If the effect is instead equitable rescission, then relief can be withheld where the interests of a *bona fide* third party purchaser intervenes. A number of responses immediately spring to mind. First, the preservation of the equitable doctrine, in and of itself, does nothing to protect third parties. If, as envisaged prior to *The Great Peace*, equity intervenes only where the common law does not, any third parties that would be prejudiced by the common law doctrine would still remain prejudiced notwithstanding the equitable jurisdiction. It is only if equity is allowed to swallow up the common law that one may

[67] In this context, litigation should not be regarded as carrying any particular negative connotations. Save in very clear-cut cases, such as *res sua* or *res extincta*, parties are unlikely to agree as to whether or not the mistake is sufficiently serious as to lead to the discharge of the contract. Discouraging litigation may lead to some mistaken parties suffering the adverse consequences of the mistake even though they have a legitimate claim to relief because they are reluctant to seek relief.

[68] As Hoffmann LJ reminded us in *William Sindall*, above n 21, at 1035: 'When [Steyn J in *Associated Japanese Bank*] speaks of the contract allocating risk "by express or implied condition precedent or otherwise" I think he includes rules of general law applicable to the contract and which, for example, provide that, in the absence of express warranty, the law is caveat emptor.' The already short list of cases on common mistake are then whittled down further when his Lordship observed that 'neither in *Grist v Bailey* [1967] Ch 532 nor in *Laurence v Lexcourt Holdings Ltd* [1978] 1 WLR 1128 did the judges who decided those cases at first instance advert to the question of contractual allocation of risk. I am not sure that the decisions would have been the same if they had.'

[69] Above n 2, at 690–91.

begin to protect innocent third parties.[70] Therefore, it is difficult to see how third parties have been prejudiced simply by the abolition of an arguably wider jurisdiction to avoid contracts, albeit one carrying a discretionary element to avoid prejudicing these self-same third parties.

Perhaps more significantly, it is important to bear in mind that common mistake operates in a very narrow range of cases, even when freed from the suffocating grip of *Bell v Lever Bros Ltd*. It simply has not been demonstrated that many (or even any) third parties are likely to be adversely affected in circumstances that are demonstrably unfair. In this respect, a few considerations must be borne in mind. The clearest examples of operative common mistakes are *res sua* and *res extincta*. But in both cases, declaring the contract void cannot possibly adversely affect any innocent third party purchasers. In the case of *res sua*, the original purchaser already had title and therefore is perfectly capable of passing on that title to an innocent third party purchaser. In the case of *res extincta*, the subject matter has ceased to exist so that the question of whether or not the original purchaser can pass title on to the innocent third party purchaser becomes moot.

Denning LJ, in inventing the equitable jurisdiction to intervene in *Solle v Butcher* had cunningly drawn upon cases of misrepresentation and unilateral mistake.[71] In so doing, Denning LJ was able to create the impression that the sympathies which tend to lie with innocent third party purchasers in the context of swindlers who impersonate others in order to obtain possession of goods on credit apply equally in cases of common mistake. However, careful consideration shows that they surely do not. In cases of mistaken identity, the fraudster typically obtains possession of the goods on credit by impersonating someone else whom the mistaken party is willing to extend credit to. The fraudster then sells the goods to an innocent third party purchaser and then either, in the usual case, disappears or, more rarely, is caught but found to be insolvent. The same pattern does not repeat itself in cases of common mistake. The original purchaser typically does not disappear because there is not the same incentive to abscond in cases of common mistake. The innocent third party purchaser can, therefore, generally locate and sue his or her counterparty, the original purchaser, should title fail to pass. If the original purchaser should prove to be insolvent, it is likewise difficult to see why our sympathies should lie with the innocent third party purchaser rather than the seller. In the context of mistaken identity, it has been suggested that it is preferable as a matter of legal policy to allocate the loss to the seller who has agreed to sell on credit rather than the purchaser. As Lord Nicholls notes in *Shogun*

[70] A Phang, 'Common Mistake in English Law: The Proposed Merger of Common Law and Equity' (1989) 9 *Legal Studies* 291.

[71] See *Solle v Butcher*, above n 2, at 690–91.

Finance Ltd v Hudson, 'it is surely fairer that the party who was actually swindled and who had an opportunity to uncover the fraud should bear the loss rather than a party who entered the picture only after the swindle had been carried out and who had none'.[72]

None of these concerns apply in the context of cases of common mistake. There is no swindler and hence no swindle to discover. A superficial reading of Lord Millett in *Shogun Finance Ltd v Hudson* appears to suggest a different reason for favouring the innocent third party purchaser. According to his Lordship, '[a]s between two innocent persons the loss is more appropriately borne by the person who takes the risks inherent in parting with his goods without receiving payment'.[73] Taken literally, and it is far from clear that this was intended by his Lordship given the context in which this passage appears, this view would apply with equal force in the context of common mistake. It is, however, a far less persuasive policy argument than that articulated by Lord Nicholls for surely the immediate response would be to ask, 'What of the risk undertaken by the innocent third party purchaser that the original purchaser had no title?' At best, it can be said that both the seller and the innocent third party purchaser assumed certain risks which have eventuated and it is difficult to see why one should be preferred over the other. The truth, therefore, is that the result of nullity is hardly as deleterious to third party rights as the picture Denning LJ painted, nor is the case for protection of such rights as may intervene, as obvious as is usually assumed.

Finally, it is suggested that if the courts consider that such third party rights are truly a legitimate concern, there is no logical reason why the common law cannot be moulded to address this concern. Cases of common mistake and frustration have for some time been regarded as closely related. It is today generally acknowledged by the courts that neither doctrine rests upon the implied intentions of the parties. Such implied intention is entirely fictitious.[74] In the context of frustration, it has been suggested that concerns over the theoretical basis of the doctrine have

[72] [2004] 1 AC 919 (HL) [82].

[73] *Ibid*, at [35].

[74] In the context of frustration: see *Davis Contractors Ltd v Fareham Urban DC* [1956] AC 696 (HL) 715 (Viscount Simonds), 720 (Lord Reid) and 728 (Lord Radcliffe). This point is perhaps best put by Lord Radcliffe who remarked that 'there is something of a logical difficulty in seeing how the parties could even impliedly have provided for something which ex hypothesi they neither expected nor foresaw' (at 728). In the context of common mistake, the Court of Appeal in *The Great Peace* (above n 1, at [73]) remarked that 'the theory of the implied term is as unrealistic when considering common mistake as when considering frustration. Where a fundamental assumption upon which an agreement is founded proves to be mistaken, it is not realistic to ask whether the parties impliedly agreed that in those circumstances the contract would not be binding.'

little practical significance.[75] This is not so. So long as the implied term theory ruled common mistake and frustration, the necessary result of an operative common mistake would be to render the contract void *ab initio* (on the basis of an implied condition precedent) whereas the inevitable result of operative frustration would be to render the contract void as from the occurrence of the frustrating event (on the basis of an implied condition subsequent). Divorced from the implied term theory, it should be open to the courts to decide that, instead of common mistake resulting in the contract being void, it could simply render the contract voidable at the instance of either party. After all, that is what the courts have suggested occurs in equity for common mistake and no one has suggested that its remedial response, as opposed to the presence of the jurisdiction, is mistaken. The remedy of rescission is certainly not unique to equity since the common law exercises precisely this remedy in the context of duress. Why then should it not be available, if deemed desirable, in the context of common mistake?

C. Remedial Flexibility

This leaves then the last concern, which must also be the most controversial—that of remedial flexibility in equity. The obvious first point, and readers must surely be feeling a sense of déjà vu by now, is surely that if remedial flexibility is desirable, why should its availability be dependent on the difficult distinction between a mistake which is 'fundamental' and one which renders the contract 'essentially different'? If remedial flexibility is desirable, then surely the status quo pre-*The Great Peace* is not satisfactory, at least not without artful manipulation of the scope of the common law doctrine by the courts. The logical argument must again be to allow equity to swallow up the common law so that there is remedial discretion in *all* cases of common mistake.[76] But before such a view can be acceded to, it must first be determined if the remedial flexibility apparently provided by equity is desirable.

As a matter of authority, its pedigree is certainly suspect. Denning LJ purported to derive his view of remedial flexibility from *Cooper v Phibbs*,[77] where the House of Lords ordered rescission on terms, arguably on the basis of common mistake.[78] However, as the Court of Appeal in *The Great Peace* observed, the House of Lords had not purported to assume a broad discretion to impose terms as Denning LJ seemed to suggest but had

[75] GH Treitel, *Frustration and Force Majeure*, 2nd edn (London, Sweet & Maxwell, 2004) 16–013–16–016.
[76] Phang, above n 70.
[77] (1867) LR 2 HL 149.
[78] Slade, above n 19.

merely imposed terms to protect the rights of parties affected by the decision.[79] The principal objection to a broad discretion to do justice between the parties is not that it is not a desirable objective, but that it is an undesirable *means* to a desirable end. Discretion without direction is an invitation to uncertainty, arbitrariness and abuse. 'Doing justice between the parties' surely cannot be sufficient direction. It is necessary then to distill the particular concerns inherent in the exercise of such discretion. Apart from concerns to protect innocent third parties,[80] three other functions of remedial discretion may be discerned from the cases dealing with common mistake and frustration. First, the remedial discretion would allow the courts to reverse any unjust enrichment on the part of either party. Secondly, it may also be used to allocate losses incurred by either or both parties. Finally, the broad discretion may be used to craft a new agreement for the parties. Each of these functions, when properly examined, will be found to be unnecessary, undesirable or otherwise less than ideal. Furthermore, that the remedial discretion has been sparingly used will not come as a surprise given how rarely the courts need to resort to the doctrine of common mistake. Even when a case of common mistake can be established, there is a general reluctance to engage in discretionary revisions of rights which would otherwise flow from rescinding the contract since such consequences are generally just and sound.[81]

The use of the remedial discretion to reverse any unjust enrichment, whilst arguably useful when Denning LJ first formulated the discretion in *Solle v Butcher*, is surely today redundant in light of developments in the law of unjust enrichment.[82]

The use of the discretion to allocate losses, whilst not featuring prominently in cases of common mistake, has sometimes found expression in cases of frustration, in the context of the limited flexibility afforded by the Law Reform (Frustrated Contracts) Act 1945.[83] However appealing it may initially appear, the use of discretion to allocate losses is a controversial proposition. The doctrinal basis of such a power is unclear[84] and this makes it difficult to determine when the loss incurred by one party should be shared by the other, and if so, how much should be thus shared. Such a

[79] *The Great Peace*, above n 1, at [105]–[108].

[80] See text accompanying nn 69–75.

[81] In *West Sussex Properties*, above n 32, the Court of Appeal refused the request of the defendant that rescission in equity be subject to a term that past overpayments in rent should be excluded. The Court of Appeal considered ordinary principles of unjust enrichment in determining that terms should not be imposed.

[82] Which first received formal judicial recognition in the House of Lords in *Lipkin Gorman v Karpnale Ltd* [1991] 2 AC 548 (HL).

[83] *Gamerco SA v ICM/Fair Warning (Agency) Ltd* [1995] 1 WLR 1226 (QBD).

[84] R Stevens, 'Three Enrichment Issues' in A Burrows and Lord Rodger (eds), *Mapping the Law: Essays in Memory of Peter Birks* (Oxford, Oxford University Press, 2006) 49, 59.

duty to share in the loss would also be, if not anathema to the predominantly adversarial view of contracting adopted by the common law, at least slightly out of place in a legal system devoid of special pre-contractual duties. If it is felt that such pre-contractual duties ought to be introduced into the common law, there seems to be no reason why this change needs to be effected in equity rather than at law, now that *The Great Peace* has decisively dissociated the doctrine of common mistake from the implied terms theory. Perhaps more importantly, any such duty to share is likely to form but a small part of a more significant doctrinal shift in contract law theory, not unlike the development of the law of unjust enrichment, and hiding and corralling such a development in the tiny corner of contract law that is common mistake would do a disservice to the law.

The most controversial use of the discretion must, however, be its use to redraft the contract for the parties. The first case where this occurred, of course, is *Solle v Butcher* itself. The facts have been earlier rehearsed and will not be repeated.[85] It will be recalled that the majority of the Court of Appeal expressed the view that both Solle and Butcher had laboured under a common fundamental mistake which, whilst not affecting their lease agreement at law, had the effect of rendering the contract voidable in equity. After inventing the remedial discretion in equity, Denning LJ opined that the circumstances were such that Butcher would only be permitted to rescind the contract in equity upon terms.[86] The terms were as follows: Butcher must be prepared to give an undertaking that he would permit Solle to be a licensee of the flat until a new lease was granted. During this interim period, Butcher must comply with the procedures which would allow him to make additions to the standard rent. Thereafter, *if* Solle made a written request for a lease, Butcher was obliged to grant him a new lease on the same terms as the rescinded lease, save in one respect. The rent for the new lease would be for *the full permitted rent, not exceeding £250 per annum*. However, Solle was neither obliged to accept the licence nor the new lease.

The second case in which the equitable discretion was exercised to impose terms by way of revising the parties' contract was *Grist v Bailey*, a decision of Goff J.[87] Somewhat simplified, the facts are as follows. The plaintiff, Frank Grist, had contracted to purchase for £850 a freehold property from the defendant, Minnie Bailey, 'subject to the existing tenancy thereof'. It transpired, however, that the existing statutory tenants had passed away so that the market value of the house was £2,250. When Grist sued for specific performance, Bailey pleaded rescission for common mistake. Goff J held that the contract was voidable in equity for common

[85] See text accompanying n 15.
[86] *Solle v Butcher*, above n 2, at 697.
[87] [1967] Ch 532.

mistake. As was the case in *Solle v Butcher*, Goff J was prepared to impose terms. However, the circumstances leading to Goff J's readiness to do so are significantly different. It appears that Goff J had not considered the imposition of terms of his own accord but because Bailey had offered to submit to a term that rescission should be on condition that she enter into a fresh contract at a proper vacant possession price. Nor is it clear that terms were actually imposed because Goff J had merely indicated that he would impose such a term *if required by Grist*.[88] This hardly amounts to imposition.

Grist v Bailey thus does not offer much support to Denning LJ's broad equitable discretion to impose terms by way of revising parties' contracts. This leaves us with only *Solle v Butcher* and it is a poor champion for the broad equitable discretion to rewrite terms. At first sight, the case looks a promising one for the imposition of a fresh contract. Solle and Butcher arrived at an agreement of the terms of their contract, and the terms were ineffective only because, by mistake, Butcher had failed to comply with the relevant procedure to have the standard rent increased. Under such circumstances, why shouldn't the courts, in rescinding the contract, simply impose the exact same terms upon the parties, *after* the relevant procedural steps have been taken? After all, is it not the case that the contract is being avoided so as to relieve the parties of the effects of their common fundamental mistake? If relieving them of their common fundamental mistake means that they would have entered into a different contract, then why shouldn't the court impose such a contract on the parties? As an objective, it is certainly hard to fault. However, its implementation is fraught with difficulties. Cases in which it is clear that the parties would have entered into a particular contract on particular terms if they had been aware of their shared mistake will be few and far between. Coupled with the fact that the doctrine of common mistake remains an exceedingly narrow one, it is questionable if such a case would reveal itself more than once in every few decades. Certainly, the curious facts of *Solle v Butcher* have not been seen again since 1949 and more than half a century has passed since. The usefulness of such discretion must be weighed against the temptation it will pose to judges to extend this discretion beyond such clear cases. Outside of such cases, the discretion to impose a contract upon the parties will be exercised *not* to relieve the parties of the consequences of their mistake but to impose upon them a contract which the court considers fair.

Solle v Butcher itself reveals further difficulties with this use of the discretion. Although it appears as if the new contract imposed upon Solle and Butcher is in all respects the same as that rescinded by the court, this is

[88] *Ibid*, at 543.

not the case. The rent under this new lease agreement is not £250 per annum but the maximum permitted rent subject to a cap of £250 per annum. Whereas this may be fair to Solle, it is hardly fair to Butcher. Suppose the parties were not mistaken and Butcher had applied for additions to the standard rent. Suppose additions up to £300 per annum were permitted, are we to conclude that he would have been content to lease the flat to Solle for £250 per annum? Not only are the *terms* of the new contract somewhat lopsided, Denning LJ further weights the conditions to rescission in favour of Solle by providing that although Butcher is obliged to grant first the licence and then the new lease to Solle, Solle is not obliged to take up either offer. In short, the terms would be imposed only upon Butcher but not Solle should Butcher choose to rescind the contract. The result, of course, is that if there is a catastrophic collapse of the rental market, Solle is free to abandon the rescinded lease agreement and take up a fresh lease elsewhere. Thus, he is protected from any upside that may follow from approvals for rent increases beyond the contractually agreed £250 per annum. If it demonstrates anything at all, *Solle v Butcher* shows that an unrestrained discretion to impose a contract upon the parties is a very dangerous thing.

Two courses seem open, therefore, so far as imposing new contracts upon parties are concerned. The courts may take the view that they should never be able to impose a contract upon the parties because the advantages that such a power would provide in the impossibly rare cases that it will be appropriately exercised are outweighed by the dangers that it poses when wrongly exercised. If, however, the courts consider that the dangers do not outweigh the benefits that such a power will bring in those few cases, then it surely behoves them to articulate the basis upon which the power will be exercised. It is difficult to imagine any other legitimate purpose such a power can serve beyond imposing upon the parties a contract which they would have made had they not been mistaken at the time of contracting. The difficulty here, of course, is determining when such exceptional circumstances exist since it would be akin to locating a needle in a haystack under the light of the blue moon. Neither option is self-evidently correct though it should be noted that with the dissociation of the common law doctrine from the implied terms theory, there is no reason why such 'discretion' should not be available at common law. The most appropriate remedy for a common mistake must, in theory at least, surely be to relieve the parties of the effects of that mistake. If it can be shown that the parties would have entered into a different contract, then it seems perfectly legitimate for the courts to impose *that contract* upon *the parties* as relief from common mistake. More often than not, however, this will not be demonstrable and the more appropriate remedy will be to declare the contract void.

IV. CONCLUSION

Common mistake after *The Great Peace* is a very narrow doctrine, but not because the courts have crafted a very narrow test. It is a narrow doctrine because in most cases where the parties are afflicted by a common fundamental mistake, the consequences will be addressed by the contract itself, either through its express or its implied terms as to the allocation of risk. It is only when the parties' agreement *and* the common law's default rules on risk allocation run out that the doctrine of common mistake operates. It is demonstrable, and it is hoped has been demonstrated in this article, that many of the concerns over *The Great Peace* are either unfounded or overstated. Furthermore, as a result of focusing on these well-rehearsed concerns, the implications for the law of common mistake as a result of the clear dissociation of the doctrine from the implied terms theory have not been fully appreciated. *The Great Peace* does not lay down a perfect doctrine of common mistake but it has clarified the law to a very large extent and has further laid the foundations for improvements. Rather than harping on the past by bemoaning the lost equitable jurisdiction, scholars and judges alike should look to build upon *The Great Peace*. If they did so, they may find something far more valuable than the flawed equitable jurisdiction that was 'lost' out at sea.

14

Contractual Mistake, Intention in Formation and Vitiation: the Oxymoron of Smith v Hughes

MINDY CHEN-WISHART*

SMITH V HUGHES is a venerable and often cited case which is as familiar as it is foundational to students' understanding of contract law.[1] Familiarity can impair clarity of vision, but repeated explanations of the case to students can also give rise to a creeping sense of unease that something is amiss. A quick recap of the case: a race horse trainer agreed to buy oats from a farmer after inspecting a sample which the trainer mistook for *old* oats. When the oats turned out to be *new* and of no use to him, the buyer refused to pay. The trial judge instructed the jury that the buyer could only win on two bases. The jury found for the buyer but did not state the basis of its decision. On appeal, the court found no evidence for the first basis stated by the trial judge; namely, that the buyer was not obliged to pay for new oats if the agreement was for old oats. The seller gave no express or implied warranty of 'oldness' and the oats delivered corresponded with the sample the buyer had inspected and agreed to buy. That left the trial judge's second basis; namely, that the seller knew that the buyer 'believed, or was under the impression, that he was contracting for old oats'.[2] The trouble was the ambiguity of this direction given by the trial judge. It did not differentiate between two types of seller knowledge, only one of which would void the contract. That is, the contract is only void if the buyer believed that the seller was *promising* the oats to be old, but not if the buyer merely *believed* the oats were old when they were not. The lack of evidence for the former (that is the seller's

* Thanks to John Cartwright, Jane Stapleton, Laura Hoyano, Kate Blackmon and participants of the 'Exploring Contract Law' symposium at the University of Western Ontario.

[1] (1871) LR 6 QB 597.

[2] *Ibid*, at 602.

knowledge of the buyer's mistake as to the promise or as to the terms) suggested to the appellate court that the jury had wrongly voided the contract for the latter (that is the seller's knowledge of the buyer's mistake as to quality). The verdict being unsafe, a new trial was ordered.

Smith v Hughes is regarded as authority for three propositions:

(i) The law determines contracting parties' intentions *objectively*.
(ii) However, if one party makes a mistake (that is he or she *subjectively* believes something different) about the *terms* of the contract, the contract is void if the other party knows of this mistake.
(iii) In contrast, if one party makes a mistake as to *fact* (that is an unwarranted background assumption) the contract is only void if the mistake is fundamental and shared by both parties.

The first proposition seems straightforward and, as we will see, has been too easily assumed without sufficient attention to its attributes. But, the initial plausibility of the second proposition turns into a growing unease on even fleeting reflection. It requires the student to accept an incoherent picture of contract law which confidently asserts the dominance of the *objective* test in ascertaining contract parties' intentions, but occasionally permits a switch to a *subjective* approach. Indeed, this is the picture painted by several distinguished commentators[3] who state that contract law recognises exceptional cases when the subjective approach trumps the objective approach to void contracts for 'mistake as to terms'. For example, the newest edition of *Treitel: The Law of Contract* states that the objective principle 'is not purely objective',[4] there being 'three exceptional situations, in which the objective principle does not apply, so that the

[3] See, eg, GH Treitel, *The Law of Contract*, 11th edn (London, Sweet & Maxwell, 2003) 307 (there are 'three exceptional situations, in which the objective principle does not apply, so that the mistake is *operative*'); J Beatson, *Anson's Law of Contract*, 28th edn (New York, Oxford University Press, 2002) 321ff (in some situations 'the law regards a contract as void though at first sight it appears perfectly valid'); AS Burrows, *A Casebook on Contract* (Oxford, Hart Publishing, 2007) 592 ('four main exceptions' to objectivity); PS Atiyah, *An Introduction to the Law of Contract*, 5th edn (Oxford, Clarendon Press, 1995) 84 (in some cases 'the law abandons the objective interpretation of the first party's intentions'); E McKendrick, *Contract Law*, 7th edn (New York, Palgrave Macmillan, 2007) 23 ('two situations in which the objective test is either displaced or modified by a test which at least on the face of it appears to place greater emphasis upon the subjective intentions of the parties'); E McKendrick, *Contract Law: Text, Cases and Materials* (Oxford, Oxford University Press, 2003) 25 (there are '[c]ases in which it has been argued that the courts have resorted to a subjective approach').

[4] E Peel, *Treitel: The Law of Contract*, 12th edn (London, Sweet & Maxwell, 2007) 1–002.

mistake is operative'.[5] This characterisation begs the questions: *why* should objectivity give way in these circumstances and *when* else might it do so?[6]

Part I begins to answer these questions by examining the justifications for and the meanings of the objective test of contractual intention. I argue that contract law's purpose in protecting the autonomy-enhancing institution of contract justifies the objective test. The test determines *whether* the parties reached agreement and *what* their agreement was for, by the way it attributes meaning to the parties' conduct. I argue that the correct version of the objective test is one that: (a) attributes meaning from the perspective of the party *observing* the conduct; (b) assumes the honesty and reasonableness of the party engaged in the conduct; and (c) takes into account the *context* of the parties' dealings. But, as I further argue, the objective approach is not omniscient. It is not all-seeing and can falter.

Part II then applies this conception of the objective test to the cases of so-called 'mistake of terms' and argues that the latter are merely examples of this objectivity properly conceived. They can only be described as an exceptional situation of the subjective trumping the objective by assuming an improper test of objectivity. Talk of 'mistake' here leads to muddle and confusion. Moreover, I argue that, on the proper test of objectivity, it is logically impossible to make the fact-finding necessary to void a contract for 'known mistake of terms'. To read *Smith v Hughes* as authority for both the objective test of intention and for voiding a contract where one knows of the other's subjective mistake of term is an oxymoron.

Part III explores the troubling proposition (iii) above. The distinction between a 'known mistake as to *term*' and a 'known mistaken assumption as to *fact*' seems paper-thin, especially when they both relate to the subject matter of the contract. In *Smith v Hughes*, Hennen J said that to void the contract, 'the jury should find not merely that the [seller] believed the [buyer] to believe that he was buying old oats, but that he believed the [buyer] to believe that he, the [seller], was contracting to sell old oats'.[7] Jack Beatson offers a clarification of this distinction. According to him, where A sells X a piece of china: the contract stands where 'X thinks that it is Dresden china. A knows that X thinks so, and knows that it is not.'[8] This is because 'X's error is one of motive alone'.[9] In contrast, the contract is void where:

> X thinks it is Dresden china, and thinks that A intends to contract to sell it as Dresden china. A knows that X thinks A is contracting to sell it as Dresden

[5] *Ibid*, at 8–048–8–052. According to Peel, these three exceptional situations are: 'mistake known to the other party', 'mistake negligently induced' and 'ambiguity'.

[6] See H Collins, *The Law of Contract*, 4th edn (London, LexisNexis, 2003) 235: 'The inadequacy of this analysis becomes plain when the textbooks offer lists of exceptions that differ from one another both in their number and character.'

[7] *Smith v Hughes*, above n 1, at 610–11.

[8] Beatson, above n 3, at 324.

[9] *Ibid*.

> china, but does not mean to, and in fact does not, offer more than china in general terms. There is no contract to sell the particular piece of china. X's error … [was] as to the nature of A's promise.[10]

While students might be prepared to accept the theoretical possibility of the distinction, they struggle with its practical application. Moreover, the justification for the different remedial thresholds set for different types of mistakes remains unclear. If a mistake as to some non-fundamental quality of the subject matter (for example that the oats are old) does not negate consent, why should a mistake as to whether the other party promised it to be old be regarded as negating consent? And this is just for starters. How should we understand the other categories of mistake (for example mistake of identity, mistake in recording the document, mistake about the nature of the document signed and so on) and their own distinctive tests for relief? The student's defensive and understandable response is to suspend disbelief, learn off by heart what is not understood and move on quickly in search of firmer ground.

The root of the trouble is the instability of the language used in this area of the law. The looseness of the key descriptive and prescriptive terms—'mistake', 'defective consent', 'objectivity' and 'void'—allows quite different problems to be thrown together resulting in muddle, confusion, obfuscation and loss of precision. To cut through this Gordian knot we need to identify the correct taxonomy of 'mistake' and 'defective consent' and stabilise the meanings of 'void' contracts and the 'objective' test of intentions. My contention is that so-called 'mistake of terms' cases are properly located in stage one of the enforceability question, namely, in contract *formation*. In contrast, 'mistake of fact' cases belong in stage two which deals with the *vitiation* of contract. This reflects the critical distinction between disappointed expectations in relation to contractual terms and those in relation to non-terms ('facts' or 'mere representations'), a distinction which is fundamental in the law of misrepresentation and breach. This classification explains why mistakes as to a contract's term only have to be known by the other party while mistakes of fact must be shared and fundamental in order to void contracts. We can also begin to taxonomise the related areas of rectification, *non est factum*, mistaken identity and misrepresentation and see the direction in which the coherent resolution of troublesome issues arising under those headings should go.

I. DEFENDING AND DEFINING OBJECTIVITY

Two approaches are generally contrasted in determining the intentions of a party: the *subjective* approach refers to a party's *actual* intention, regardless

[10] *Ibid*.

of what he or she appears to intend from his or her conduct; and, the *objective* approach refers to what a *reasonable* person would interpret as a party's intention from his or her conduct in all the circumstances. *Smith v Hughes* affirmed the dominance of the objective approach. In Lord Blackburn's familiar dicta:

> 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.[11]

The often stated requirement of *consensus ad idem* must be filtered through this lens of objectivity. The logical corollary is that genuine agreement is unnecessary; the mere *appearance* of agreement is enough. Any deviation in a party's actual intention from this objective reference point is logically and legally irrelevant.

A. The Justification for Objectivity

The subjective–objective distinction is real. A party's objective intention may, but need not, correspond with his or her subjective intention. A party's conduct may signify consent to a contract when he or she did not *really* consent or when he or she believed that they were consenting to different terms. Parties *can*, without any dishonesty, misrepresent their own meaning and misinterpret the meaning of others. Why is the objective approach preferred when its effect is to bind parties to contracts that they did not intend?

Even if contract law is really concerned about the parties' subjective intentions, the parties' objective intentions may be the best proxy for two pragmatic reasons. The first is one of accessibility. Since 'the intent of a man cannot be tried, for the Devil himself knows not the intent of a man',[12] the objective approach overcomes the evidential difficulties in determining what was *really* in a party's mind at the relevant time. The second goes to avoidance of fraud. To determine a person's intention simply by reference to what he or she asserts was his or her undisclosed subjective intention (which may otherwise be impossible to access) is to invite dishonesty and chaos. Once a conflict has arisen, a person's incentive to distort the truth in favour of his own self-interest (even subconsciously) should disqualify such evidence.

[11] *Smith v Hughes*, above n 1, at 607.
[12] *Anon* (1478) YB 17 Edw 4, Pasch fo 1, pl 2.

The objective approach, however, is also justifiable on its own terms, for three further reasons. The first looks at certainty and the protection of reasonable expectations. Great disruption would ensue if I could escape the liability I appear to have assumed to you by simply asserting that I *really* meant something different. Even if my real intention is objectively provable from facts unknown to you, it would be unjust to prioritise it over your honest and reasonable, but *different*, interpretation of my outward conduct. One of contract law's main functions is to facilitate the security of transactions and so enable people to plan and shape their lives on the basis of an apparently enforceable contract. This function would be hopelessly undermined if legal significance were to attach to the claim that: 'When I say "white" I mean "black."' The objective approach allows parties to know in advance how their own conduct will be interpreted and how they are entitled to interpret the conduct of others. This overlaps with and is reinforced by the next justification.

The next justification is based on the protection of the autonomy-enhancing institution of contracting. The purpose of contract law is to protect the *practice* of undertaking voluntary obligations because it enables parties to act autonomously, to make their own arrangements, to shape their own lives.[13] The objective test of intentions is one of the rules of engagement necessary to protect the integrity of the contracting process and to prevent its abuse.[14] Holding parties to the objective standard: (a) prevents them from reneging on their undertakings; (b) gives them strong incentives to take care not to misrepresent their own intentions (even innocently) nor to misinterpret the intentions of others; and (c) extends the practice beyond ongoing relationships where it would otherwise not exist: 'But for the support of the law, contracts between complete strangers would not be as numerous and common as they are.'[15]

Finally, objectivity is intrinsic to contracting. Making a contract is essentially an exercise in the *communication* of choice, and communication is impossible without objectivity.[16] We have to suspend our own meaning, enter imaginatively into the other's world and ask: 'What meaning do *they* think they're conveying?' and 'What will *they* think I am meaning?'[17] As May LJ said in *Ove Arup v Mirant Asia Pacific Construction Ltd*: 'Subjective intention or understanding, unaccompanied by some overt

[13] J Raz, 'Promises in Morality and Law' (1982) 95 *Harvard Law Review* 916, 933.
[14] *Ibid*, at 928–38.
[15] *Ibid*, at 934.
[16] T Endicott, 'Objectivity, Subjectivity, and Incomplete Agreements' in J Horder (ed), *Oxford Essays in Jurisprudence (4th Series)* (Oxford, Oxford University Press, 2000) 151; D Goddard 'The Myth of Subjectivity' (1987) 7 *Legal Studies* 263.
[17] A Kronman, 'Paternalism and the Law of Contracts' (1983) 92 *Yale Law Journal* 763.

objectively ascertainable expression of that intention or understanding, is not relevant.'[18] Similarly, Ernest Weinrib explains that objectivity is intrinsic to private law since:

> Interaction between free wills engages the external aspect of practical reason … On stepping into a world of interaction, the freely willing actor establishes a presence there through acts that have an externally recognizable nature. Purely mental imaginings and reservations, however real they are to the actor or however serious the consequences to which they might in due course lead, have no status in this world of interaction. Thus criminal wrongdoing requires an *actus reus*; contract cannot be held hostage to the vagaries of a private intention; and the claim to property must involve some act in the world of appearances … The external nature of action implies a world of shared social meanings. Only within such a world can juridical acts by each of the parties be interpreted from a perspective common to both and thus have significance as *external* acts.[19]

Intention is wholly dependent on manifestations interpreted in a context of shared meaning. The existence and extent of contractual obligations are determined by the signs made—the moves in the language game being played. Any legal concern with undisclosed intention is senseless: it contradicts the very idea of contract as an agreement between parties who convey and receive meaning.

B. Objectivity and Voluntariness

The objective test only bites when it yields a result that trumps a party's actual and different intention. This fact is highlighted by those advancing hybrid principles of contractual liability based on reliance, unjust enrichment, fairness, and public policy[20] over one based on the parties' voluntary exercise of autonomy. Objectivity undoubtedly undermines the view of contract law based on the importance of promise-keeping.[21] But, it is consistent with the view that one of contract law's primary functions is to protect the facilitative institution of contracting.[22] That is, 'to protect both

[18] [2004] BLR 49 (CA) [62].

[19] E Weinrib, *The Idea of Private Law* (Cambridge, Harvard University Press, 1995) 104.

[20] See, eg, G Gilmore, *The Death of Contract* (Columbus, Ohio State University Press, 1974); P Atiyah, *The Rise and Fall of Freedom of Contract* (Oxford, Clarendon Press, 1979); P Atiyah, *Essays on Contract* (Oxford, Clarendon Press, 1986).

[21] C Fried, *Contract As Promise* (Cambridge, Harvard University Press, 1981) 61: 'the court imagines that it is respecting the will of the parties by asking what somebody else, say the ordinary person, would have intended by such words of agreement … it palpably involves imposing an external standard on the parties … clearest in cases of what is called unilateral mistake.'

[22] Raz, above n 13, at 937: 'To enforce voluntary obligations is to enforce morality through the legal imposition of duties on individuals. In this respect it does not differ from the legal proscription of pornography.' For a similar view, see Endicott, above n 16; Goddard, above n 16; H Sheinman, 'Contractual Liability and Voluntary Undertaking' (2000) 20 *OJLS* 205.

the practice of undertaking voluntary obligations and the individuals who rely on that practice… One protects the practice of undertaking voluntary obligations by preventing its erosion—by making good any harm caused by its use or abuse.'[23] The objective principle prevents people from abusing the *practice* of contracting by making it appear that they have agreed to obligations when they have not. It is consistent with the harm principle[24] which holds that the only proper purpose for imposing legal obligations on individuals is to prevent harm. But unlike reliance theorists who focus on harm to the individual, Joseph Raz extends the notion to include 'institutional harm'. Liability is not imposed to protect voluntariness on the *individual* level (except at the very high threshold level of *non est factum*),[25] but in order to protect the *practice* of undertaking voluntary obligations; to preserve its appeal and utility.

Thus, voluntariness, albeit on an institutional level, remains the distinctive touchstone of contractual liability. The objective test of contract formation is not an embarrassment to the view that the purpose of contract law is to support the autonomy-enhancing practice of undertaking voluntary obligations. Paradoxically, it is in order to protect the practice of contracting from debasement that the law recognises the validity of contracts that are not voluntary obligations.[26]

C. Defining Objectivity

A contract is comprised of the coincidence of the parties' voluntary intentions to be bound by certain terms. How do we know what each party intended? How is the meaning of another person constructed, communicated, interpreted, or accessed?[27] Securing the priority of objectivity over subjectivity is only the starting point. Three further questions arise: (i) Objectivity from whose perspective? (ii) Objectivity on what standard? (iii) Objectivity on what evidence?

(i) Objectivity from Whose Perspective?

William Howarth sets out three perspectives from which a party's intention has been assessed by the courts.[28] The first is detached objectivity,

[23] Raz, *ibid*, at 933.

[24] JS Mill, 'On Liberty' in *Three Essays: On Liberty, Representative Government, The Subjection Of Women* (London, Oxford University Press, 1975).

[25] See text accompanying nn 106–15.

[26] Sheinman, above n 22; Raz, above n 13, at 935.

[27] See further C Dalton, 'An Essay in the Deconstruction of Contract Doctrine' (1985) 94 *Yale Law Journal* 997, 1039–65; J Finnis, 'The Priority of Persons' in J Horder (ed), *Oxford Essays in Jurisprudence (4th series)* (Oxford, Oxford University Press, 2000) 1, 11–13.

[28] W Howarth, 'The Meaning of Objectivity in Contract' (1984) 100 *LQR* 265.

sometimes called 'fly on the wall' objectivity. This perspective adopts the viewpoint of a reasonable person independent of that of either contract party. The second perspective is 'promisor' objectivity, which adopts the viewpoint of the reasonable person seeking to avoid the contract (that is the defendant). The third perspective is that of 'promisee' objectivity, which adopts the viewpoint of the reasonable person seeking to enforce the contract (that is the claimant). The terminology of 'promisor' and 'promisee' objectivity is problematic because they depend entirely on the relatively insignificant fact of which party initiates proceedings; indeed, in respect of the term in contention, either the *actual* promisor or promisee could be the 'promisor' (that is the defendant) or 'promisee' (that is the claimant) on Howarth's scheme. Moreover, even if we use 'promisor' and 'promisee' in their natural sense, each party is *both* in a bilateral contract: '[h]e is a promisor with regard to what he undertakes to perform and a promisee with regard to what he is entitled to receive'.[29]

In light of these difficulties, the more transparent distinction is between 'actor objectivity' (what a reasonable person in the *actor's* position would mean by her conduct), and 'observer objectivity' (what a reasonable person in the *observer's* position would interpret the actor's conduct as meaning). Both interpret the actor's conduct, but they may differ because the factual matrix informing each party's perspective may differ; facts affecting 'actor objectivity' may not be apparent to the observer and vice versa. For example, the observer (seller) may reasonably interpret the actor's (buyer's) conduct as meaning 'oats' when the actor has provable reasons, unknown to the observer, for reasonably intending to mean 'old oats'. Which perspective—the detached person's, the actor's or the observer's—should be adopted?

Detached objectivity is implied by Lord Blackburn's appeal to what 'a reasonable man would believe'.[30] It is strongly supported by various dicta of Lord Denning[31] and some academic writing.[32] This approach could promote certainty and protect third parties who rely on the apparent meaning of contractual documents. But it does not address the main

[29] J Vorster, 'A Comment on the Meaning of Objectivity in Contract' (1987) 103 *LQR* 274, 276–78.

[30] *Smith v Hughes*, above n 1, at 607.

[31] *Solle v Butcher* [1950] 1 KB 671 (CA) 691: '[O]nce the parties, whatever their inmost states of mind, have to all outward appearances agreed with sufficient certainty in the same terms on the same subject matter … [n]either party can rely on his own mistake to say it was a nullity from the beginning, no matter that it was a mistake which to his mind was fundamental, and no matter that the other party knew that he was under a mistake.' See also *Frederick E Rose (London) Ltd v William H Pim Junior and Co Ltd* [1953] 2 QB 450 (CA) 461 [*Rose v Pim*].

[32] See, eg, M Furmston, *Cheshire Fifoot & Furmston's Law of Contract*, 15th edn (Oxford, Oxford University Press, 2007) 306: 'the question is not what the parties had in their minds, but what reasonable third parties would infer from their words or conduct'.

justifications for objectivity (protecting parties' reasonable expectations and reliance, and the practice of undertaking voluntary obligations).[33] It may even facilitate, rather than prevent, fraud; a party could insist on the dictionary meaning of the words used even when he or she knows that that was not the agreement. The exceptions to the parole evidence rule tell against it, most notably the doctrine of *non est factum*.[34] As John Spencer observes, while '[i]t may be acceptable for the law occasionally to force upon *one* of the parties an agreement he did not want . . . surely there is something wrong with a theory which forces upon *both* of the parties an agreement which *neither* of them wants'.[35] Actor objectivity does not encounter this objection and moreover, is consistent with the first two justifications for objectivity (accessibility and avoidance of fraud); the question would be one of the availability of evidence. However, it falls short in protecting the observer's reliance and expectations and, consequently, the institution of contracting; it is also inconsistent with the nature of communication as the conveying of meaning *to* the observer.

It is observer objectivity which has the weight of authority[36] and ticks *all* the justifications for objectivity. It also receives reinforcement from an unlikely source. The doctrine of consideration justifies the observer, who has 'paid' for the promised performance, in adopting a reasonable view of the actor–promisor's conduct from his or her (the observer's) perspective. The actor who accepts 'payment' is bound by the observer's reasonable interpretation of his or her conduct. Actors cannot rely on their different meaning, even if they can prove its reasonableness from their perspective.[37]

[33] Atiyah, *Rise and Fall*, above n 20, at 663; JR Spencer, 'Signature, Consent, and the Rule in *L'Estrange v Graucob*' [1973] *CLJ* 104, 110, 112.

[34] See text accompanying nn 106–15.

[35] Spencer, above n 33, at 113.

[36] See *André & Cie SA v Marine Transocean Ltd (The Splendid Sun)* [1981] 1 QB 694 (CA) as explained in *Paal Wilson & Co v Partenreederei Hannah Blumenthal (The Hannah Blumenthal)* [1983] 1 AC 854 (HL) 865 [*The Hannah Blumenthal*]; *Chaloner v Bower* (1984) 269 EG 725 (CA); *Tankreederei Ahrenkeil GmbH v Frahuil SA (The Multibank Holsatia)* [1988] 2 Lloyd's Rep 486 (QBD) 493; *OT Africa Line Ltd v Vickers Plc* [1996] 1 Lloyd's Rep 700 (QBD Comm) [*OT Africa Line*]; *Freeman v Cooke* (1848) 2 Exch 654, 154 ER 652, 653; Vorster, above n 29, at 278, 283.

[37] *Centrovincial Estates plc v Merchant Investors Assurance Co Ltd* [1983] Com LR 158 (CA) 24 [*Centrovincial Estates*]; *OT Africa Line*, *ibid*, at 703. This is the case unless the parties are genuinely at cross-purposes as to the subject matter of the contract and the latter is so ambiguously expressed that the court cannot determine which version is more likely, as in *Raffles v Wichelhaus* (1864) 2 Hurl & C 906, 159 ER 375 (discussed below, in text accompanying nn 101–5). And see, R Stevens, 'Objectivity, Mistake and the Parole Evidence Rule' in A Burrows and E Peel (eds), *Contract Terms* (Oxford, Oxford University Press, 2007) 101, 102–103.

The contrast is with bare promises contained in a deed[38] where actor-objectivity has priority. Millett J explains in *Gibbon v Mitchell* that the deed will be set aside for mistake 'if the court is satisfied that the disponer did not intend the transaction to have the effect which it did'.[39]

(ii) Objectivity on What Standard?

Consistent with the aim of facilitating the practice of contracting and preventing its abuse, the objective approach is infused with a bias in favour of the just and reasonable interpretation, and against the unjust and unreasonable.[40] It treats a contract party as an honest and reasonable person who will not take an unjust view of the other party's intentions, nor give a dishonest or misleading view of his or her own intentions. The same idea underlies Hugh Collins' reference to the need for 'clean hands' and his identification of a 'duty to negotiate with care' and to 'refrain from unconscionable conduct'.[41] This view is consistent with John Finnis' 'conversational' model of interpretation, which prioritises the observer's interpretation, when it is subject to the qualifier that:

> [A] properly juridical interpretation will not be as ready to consider authoritative an unjust as it will a just meaning. Thus it differs from sensible conversationalists, who like good historians are quick to detect, and not too ready to overlook, interlocutors' perhaps vicious purposes and deficiencies of personal character.[42]

(iii) Objectivity on What Evidence?

How much evidence should be taken into account in determining how a just and reasonable observer would have interpreted the actor's conduct? Two versions of objectivity can be detected lying at opposite ends of a spectrum: the traditional formal objectivity and the more recent contextual objectivity. Formal objectivity is closely related to detached objectivity; it severely limits the evidence to be taken into account and prioritises them according to a fairly strict hierarchy of probative value. Thus, s*igned final writing* contained in a contractual document is the best evidence of intention (hence the signature rule, the parole evidence rule and the general effectiveness of entire agreement clauses). It is superior to *unsigned final*

[38] *Lady Hood of Avalon v Mackinnon* [1909] 1 Ch 476. The claimant settled £8,600 on both her daughters having forgotten that she had previously settled an even larger sum on the elder daughter. Her gift was set aside.

[39] [1990] 1 WLR 1304 (Ch D) 1309.

[40] Raz, above n 13, at 934–5: 'because the predominant purpose of contract law is to support existing moral practices, one might expect that the conditions of the validity of contracts will reflect common moral conceptions … to mirror common moral views'.

[41] Collins, above n 6, at 234–7.

[42] Finnis, above n 27, at 13.

writing contained in a contractual document (the 'ticket cases') or contracts formed by exchange of writing. This in turn is superior to *speech*, which is superior to *non-verbal conduct* (a nod, a wink, contractual performance), which, finally, is superior to *silence or omissions*.[43]

Formal objectivity represents a relatively depersonalised interpretation of *conduct* as opposed to a genuine search for the meaning of *the person* engaged in the conduct.[44] The temptation to formal and detached objectivity is strongest where the contract is evidenced in some document as was the case in *L'Estrange v F Graucob Ltd*[45] and *Butler Machine Tool Co v Ex-Cell-O Corp*.[46] These cases are criticised for their unrealistic analysis of the facts because the courts prioritised the signatures over other contradictory evidence of the signer's intention, which was apparent to the observer.[47] Lord Devlin observed that the signature rule, particularly in the context of standard form contracts, is premised on a 'world of make-believe which the law has created'.[48] Signatures are still treated as binding the signatory almost absolutely, as if 'some kind of magic operated to take the contract out of the usual rules that govern the formation of contracts'.[49]

Formal objectivity continues to hold sway, particularly with commercial contracts where certainty is rightly prized and a signature can reasonably be taken to manifest intention to consent to its contents. But a more expansive contextual objectivity is growing in influence, especially in non-commercial contexts. The Canadian decision of *Tilden Rent-a-Car Co v Clendenning* applied it to a hurried consumer transaction with finely printed, unexpected and harsh clauses.[50] The Ontario Court of Appeal held that 'the signature by itself does not truly represent an acquiescence to

[43] Silence generally does not constitute acceptance even if it was clearly intended: see *Felthouse v Bindley* (1862) 6 LT 157 (CP); but inaction or silence may amount to intention to abandon one's claims if accompanied by reliance: see *The Hannah Blumenthal*, above n 36, at 865; *Allied Marine Transport Ltd v Vale Do Rio Doce Navegacao SA (The Leonidas D)* [1985] 1 WLR 925 (CA).

[44] See Dalton, above n 27, at 1039–65; A De Moor, 'Intention in the Law of Contract, Elusive or Illusory?' (1990) 106 *LQR* 632, 635–55; Goddard, above n 16.

[45] [1934] 2 KB 394 (Div Ct). The court upheld a signed document excluding all the seller's liability although the seller knew that the buyer did not read the 'regrettably small print'.

[46] [1979] 1 WLR 401 (CA). The facts were as follows. B offered to sell a machine on its terms *including* a price variation clause, but E placed an order on different terms *excluding* price variation. The latter prevailed because B had signed and returned the tear-off acknowledgement slip on E's order form stating 'we [B] accept your order on the terms and conditions stated thereon' (*ibid*, at 403). The court ignored B's covering letter sent with the tear-off slip insisting on *their* original terms (ie including the price variation clause).

[47] See, eg, Spencer, above n 33, at 121–2.

[48] *McCutcheon v David MacBrayne Ltd* [1964] 1 WLR 125 (HL) 133.

[49] Spencer, above n 33, at 117.

[50] (1978) 83 DLR (3d) 400 (Ont CA). Thus, the hirer of a car was not bound by a clause making him liable for damage to the car in a wide variety of circumstances going far beyond what he would reasonably expect since: (a) he had paid extra for additional insurance coverage; (b) it was plain that he did not read nor was expected to read the fine print; and (c) the speed with which the transaction was completed was one of the attractions of the service being offered.

unusual and onerous terms which are inconsistent with the true object of the contract' unless the business had taken reasonable steps to draw the consumer's attention to these terms.[51] The court cited Stephen Waddams:

> One who signs a written document cannot complain if the other party reasonably relies on the signature as a manifestation of assent to the contents, or ascribes to the words he used their reasonable meaning. But the other side of the same coin is that only a reasonable expectation will be protected. If the party seeking to enforce the document knew or had reason to know of the other's mistake the document should not be enforced.[52]

In English law, the general shift away from literalist interpretations of contractual documents towards a more contextual interpretation[53] culminates in Lord Hoffmann's judgment in *Investors Compensation Scheme Ltd v West Bromwich Building Society*.[54] Accordingly, the court should place itself in the same factual matrix as the parties and to take into account absolutely everything[55] reasonably available to them which would have affected the way that they interpret the contractual documents and, logically, any other manifestations of intent.[56] This approach largely discards the 'old intellectual baggage of "legal" interpretation' replacing it with 'the common sense principles by which any serious utterances would be interpreted in ordinary life'.[57] The search is for the meaning of the *person*, and not merely of the *conduct*. Consistently with this view, Finnis reasons that since law is 'for the sake of persons and its rules are fundamentally relationships between persons' the intent of persons should be treated as if it really matters.[58] Raz opposes a philosophical individualism which fails to recognise that an act has different normative implications depending on its social context.[59] Similarly, Kronman argues against

[51] *Ibid*, at 407.

[52] *Ibid*, at 405, citing SM Waddams, *The Law of Contracts* (Toronto, Canada Law Book, 1977) 191.

[53] See, eg, *Prenn v Simmonds* [1971] 1 WLR 1381 (HL) 1383–4 (Lord Wilberforce): 'In order for the agreement … to be understood, it must be placed in its context. The time has long passed when agreements … were isolated from the matrix of facts … and interpreted purely on internal linguistic considerations. We must inquire beyond the language and see what the circumstances were with reference to which the words were used. … English law [was not] left behind in some island of literal interpretation.'

[54] [1998] 1 WLR 896 (HL) 912–13 [*West Bromwich*]. See also *Lord Napier and Ettrick v RF Kershaw* [1999] 1 WLR 756 (HL) 763 (Lord Steyn): 'Loyalty to the text of a commercial contract, instrument or document read in its contextual setting is the paramount principle of interpretation … [T]he court ought generally to favour a commercially sensible construction [because the] commercial construction is likely to give effect to the intention of the parties.'

[55] See criticisms of this approach raised by C Staughton, 'How do the Courts Interpret Commercial Contracts?' [1999] *CLJ* 303.

[56] See Lord Nicholls, 'My Kingdom for a Horse: the Meaning of Words' (2005) 121 *LQR* 577 criticising the rules excluding evidence of previous negotiations and subsequent conduct.

[57] *West Bromwich*, above n 54, at 912.

[58] Finnis, above n 27, at 11–13.

[59] Raz, above n 13, at 931–2.

a clinical and impoverished approach which treats persons as denatured, disembodied egos, in favour of a context-infused richness of meaning.[60]

The contextual approach is exemplified by *Mannai Investment Co Ltd v Eagle Star Life Assurance Co* where the tenant wrote '12th' when his 'evident intention' was '13th'.[61] Lord Hoffmann explained that words do not in themselves refer to anything; it is people who use words to refer to things and it 'is a matter of constant experience that people can convey their meaning unambiguously although they have used the wrong words'.[62] His Lordship added:

> It is of course true that the law is not concerned with the speaker's subjective intentions. But the notion that the law's concern is therefore with the 'meaning of his words' conceals an important ambiguity. The ambiguity lies in a failure to distinguish between the meanings of words and the question of what would be understood as the *meaning of a person* who uses words . . . This involves examining not only the words and the grammar but the background as well. So, for example, in *Doe d Cox v Roe* … the landlord of a public house in Limehouse gave notice to quit 'the premises which you hold of me . . . commonly called . . . The Waterman's Arms'… [T]here were no such premises in the parish of Limehouse. But the tenant did hold premises of the landlord called The Bricklayer's Arms . . . The meaning was objectively clear to a reasonable recipient, even though the landlord had used the wrong name . . . There was no need to resort to subjective meaning.[63]

In the decision of the Court of Appeals of New York in *Utica City National Bank v Gunn*, Cardozo J agreed that the meaning of words 'is not always the meaning of the parties'.[64] In that case, Gunn (G) and others gave a guarantee of $115,000 'on the consideration of' Utica City National Banks' (UCNB) 'loans and discounts.' Cardozo acknowledged that: 'This looks to the future. It excludes the past,' so that the words construed alone would support G's refusal to pay absent a new loan from UCNB. Yet, evidence of the genesis and aim of the guarantee showed that both parties were well aware that it was to secure an *existing* loan of that precise amount to the principal debtor. Cardozo J therefore interpreted 'loans and discounts' to include 'renewals' consistent with the 'efficacy and purpose' of the transaction. He preferred to 'give a new shade of meaning to a word than to give no meaning to a whole transaction'.

[60] Kronman, above n 17.

[61] [1997] AC 749 (HL). The tenant had written 'January 12th' in the notice to terminate the lease when the lease stipulated that he only had power to do this on its anniversary which was January 13th. The House of Lords held that the notice was effective to determine the lease on January 13th.

[62] *Ibid*, at 774.

[63] *Ibid*, at 775.

[64] 222 NY 204 (1918) 208.

The equitable doctrine of rectification is further support for the priority of observer-contextual objectivity over detached-formal objectivity. The parole evidence rule barring appeal to evidence of intention outside the contractual document does not apply where rectification is sought since the claim is precisely that the contractual document does *not* reflect the parties' agreement as the coincidence of each party's evident intention.[65] Moreover, where the parties agree on the meaning of a particular phrase used in the contractual document, the contract can be rectified to make clear that the phrase bears the meaning agreed.[66] Rectification is also allowed where (a) one party knows of the other party's mistake about the contents of the contractual document *and* of the mistaken party's real intentions,[67] (b) fails to draw the mistaken party's attention to the mistake; and, (c) the mistake benefits the unmistaken party or prejudices the mistaken party.[68] Logically, the case for rectification should be no weaker if one party's misrepresentation has induced that shared belief in the meaning of words used in the document, even if innocent.[69] In *Curtis v Chemical Cleaning & Dyeing Co Ltd*, Curtis (C) took a wedding dress to Chemical Cleaning & Dyeing (CCD) for cleaning and was asked to sign a 'receipt' exempting CCD from liability for 'any damage howsoever arising'.[70] C did so after being assured that CCD's exemption was only in relation to any damage to beads or sequins on the dress. The dress was returned badly stained. The court concluded that CCD was not allowed to rely on the wide exclusion clause; it could only rely on the exemption to the limited extent represented.

The crucial distinction is *not* between subjectivity and objectivity, since subjectivity is simply irrelevant; but rather between detached-formal objectivity on the one hand, and observer-contextual objectivity on the other. The former might necessitate resort to the notion of exceptional 'subjectivity' to explain the outcome of cases where contracts are voided for 'mistakes of terms', but this is unnecessary on the more expansive observer-contextual objectivity.[71]

[65] *Lovell and Christmas Ltd v Wall* (1911) 104 LT 85 (CA) [*Lovell and Christmas*]. Even the presence of an 'entire agreement' clause in the contract does not prevent rectification: see *JJ Huber (Investment) Ltd v Private DIY Co Ltd* [1995] NPC 102 (Ch D) [*JJ Huber*].

[66] *London Weekend Television v Paris and Griffith* (1969) 113 SJ 222 (HC). See also *Re Butlin's Settlement Trusts* [1976] Ch 251.

[67] Actual knowledge is required, but this includes wilfully shutting one's eyes to the obvious, or wilfully and recklessly failing to make such inquiries as an honest and reasonable person would have made. The doctrine also applies to one who has made 'false and misleading statements' to divert the other from discovering the mistake, see *Commission for the New Towns v Cooper* [1995] 1 Ch 259 (CA) 280 [*New Towns v Cooper*].

[68] *Thomas Bates & Sons Ltd v Wyndhams Ltd* [1981] 1 WLR 505 (CA) 515–16 [*Bates v Wyndhams*], 520–21; *New Towns v Cooper*, *ibid*.

[69] See to the contrary *Rose v Pim*, above n 31.

[70] [1951] 1 KB 805 (CA).

[71] See, eg, the examples in n 3 above.

II. APPLICATIONS OF OBJECTIVITY IN CONTRACT FORMATION

The cases often cited as exceptional resorts to subjectivity (allowing contracts to be voided for mistake as to terms) rest on the erroneous reference point of detached-formal objectivity. In fact, they are straight applications of *observer-contextual* objectivity; talk of 'knowing of the other party's mistake' is unnecessary and logically impossible.

A. Was there an Objective Agreement?

According to *Smith v Hughes* the components of an operative mistake of terms are:

(i) There is an objectively determinable agreement (A1).
(ii) The actor mistakenly believes the agreement is not for A1 but rather for a different agreement (A2).
(iii) The actor *also* mistakenly believes the observer to be *agreeing* A2.
(iv) The observer knows of the actor's mistakes in (ii) and (iii).

My argument is that once the objective agreement A1 is found via observer-contextual objectivity, the observer (in *Smith v Hughes*, the seller) cannot logically know of the actor's (in *Smith v Hughes*, the buyer) intention to agree to A2. The observer cannot honestly and reasonably believe that the actor intends to agree to A1 (implicit in finding an agreement for A1) *and* honestly and reasonably believe that he or she intends to agree to the inconsistent A2. While, in real life, it may be accepted that we can simultaneously hold contradictory beliefs, that can hardly be, and is not, the approach of the law which must necessarily come down on one side or the other. It is worth taking time over the proof.

(i) The Objective Point of Reference

Was the contract for 'oats' as the seller alleged, or 'old oats,' as the buyer alleged? This question breaks down into two: (a) What did each party's conduct, in the factual matrix of the case, honestly and reasonably lead the other to believe about his or her intention? (b) Did the parties' objective (ostensible) intentions coincide? The court's approach and conclusion was entirely consistent with observer-contextual objectivity. The agreement was for 'oats' period; the seller gave no express or implied warranty of 'oldness' and the buyer offered a price after inspecting a sample of oats which corresponded with what was later delivered.[72] Conversely, the seller had no

[72] *Smith v Hughes*, above n 1, at 603, 605 (Cockburn CJ), 607 (Blackburn J), 609 (Hennen J).

reason to know of the buyer's intention to buy *old* oats; the buyer claimed that trainers only used old oats as a rule, but the buyer had since sold new oats to a trainer;[73] and while the price was high for new oats, oats were then very scarce and expensive.

(ii) Buyer's Mistaken Assumption about the Subject Matter of the Contract

The buyer's unilateral mistaken assumption that the oats he was buying were old is legally irrelevant in itself.[74] Even the seller's knowledge of the buyer's mistake does not change this, absent a misrepresentation or a general obligation of disclosure. Passive acquiescence in another's self-deception does not void the contract.[75] The market system rewards research and knowledge so that a knowledgeable party must generally be allowed to take advantage of a less knowledgeable (mistaken) party. Cockburn CJ gave two examples of this phenomenon: a sale of land which the buyer knows that the seller is ignorant of the existence of a mine under it,[76] and the sale of a horse 'as is', where the seller knows that the buyer mistakenly believes the horse to be sound. Both contracts are binding.[77]

(iii) The Buyer's Mistake as to Terms

Although the seller gave the buyer no reason to believe that he was *promising* old oats, the buyer may have assumed this to be true on some independent basis, such as from a third party's misrepresentation.[78] While the buyer 'believing he has purchased old oats' as opposed to merely 'believing the oats he purchased were old' will, in practice, be difficult to distinguish, it is a distinction routinely made in the context of liability for misrepresentation and breach.[79]

(iv) The Seller Knows of the Buyer's Mistake of Terms

Here is the nub. Even if we find that the buyer mistook the seller's offer, it will be impossible to find that the seller knew of it. No rational party would concede such knowledge when knowledge of the other's mistake as

[73] *Ibid*, at 602.
[74] See the discussion of operative mistaken assumption of fact at text accompanying nn 146–50.
[75] *Smith v Hughes*, above n 1, at 603.
[76] *Ibid*, at 604.
[77] *Ibid*, at 606. See also, *ibid*, at 607 (Blackburn J).
[78] There may be an action against the third party for fraud or negligent misrepresentation.
[79] See text accompanying nn 136–8.

to *subject matter* can be admitted without cost (as in (ii)). Absent such an admission, the question is whether that knowledge can be inferred from the seller's own conduct or on an honest and reasonable interpretation of the buyer's conduct. The crucial point is that this is just another way of asking the *same* question as (i), namely, 'what is the objective contract?' The *same* body of evidence is interrogated, but a contradictory answer to (i) is ostensibly required to void the contract. In fact, four answers are possible, none of which necessitates or indeed is susceptible of the description 'known mistake of term'.

First, if the buyer's conduct is honestly and reasonably interpreted by the seller as consenting to buy 'oats' and vice versa[80] then *that* is the objectively determined contract at (i). The seller cannot, at the same time, honestly and reasonably believe the buyer to be consenting to (that the seller was promising) 'old oats'. The seller cannot simultaneously believe the buyer to be consenting to two contradictory things.

Second, if the buyer's conduct is honestly and reasonably interpreted by the seller as promising 'old oats', and vice versa, then *that* is the objectively determined contract at (i). The objective reference point has simply moved. As Hannen J said:

> If … the [seller] knew that the [buyer], in dealing with him for oats, did so on the assumption that the [seller] was contracting to sell him old oats, he was aware that the [buyer] apprehended the contract in a different sense to that in which he meant it, and he is thereby deprived of the right to insist that the [buyer] shall be bound by that which was only the apparent, and not the *real bargain*.[81]

On this scenario, the 'real bargain' must be for old oats. Not only can the buyer refuse payment for the new oats, he can, in principle, enforce the contract for old oats. Although this is a theoretical possibility, there was no basis for such a conclusion on the facts of *Smith v Hughes*.[82]

Third, if the seller's conduct merely induced the buyer's mistake of fact that the oats were old (that is it was not a term promising old oats), the contract is voidable for misrepresentation.[83] Fourth and finally, if the seller

[80] The seller's conduct is honestly and reasonably interpreted by the buyer as consenting to buy 'oats'.

[81] *Smith v Hughes*, above n 1, at 610 (emphasis added).

[82] *Ibid*, at 611 (Hennen J): 'It may be assumed that the [buyer] believed the oats were old, and it may be suspected that the [seller] thought he so believed, but the only evidence from which it can be inferred that the [seller] believed that the [buyer] thought that the [seller] was making it a term of the contract that the oats were old is that the [buyer] was a trainer, and that trainers, as a rule, use old oats; and that the price given was high for new oats, and more than a prudent man would have given.' However, the seller was ignorant of trainers' buying habits and had subsequently sold new oats to a trainer; oats were also scarce at the time and therefore commanded higher prices than had traditionally been the case.

[83] *Ibid*, at 605 (Cockburn CJ): 'If, indeed, the buyer instead of acting on his own opinion, had asked the question whether the oats were old or new, or had said anything which intimated his understanding that the seller was selling the oats as old oats, the case would

neither promised nor represented the oats were 'old', but still had reason to know that the buyer did not intend to contract on his terms (simply 'oats') but without knowing what the buyer *was* intending to contract for, there would simply be no agreement between the parties.[84]

Where parties are at cross purposes on the *terms* of the contract, the objective test of intention, properly understood will yield an answer to all cases, supplemented by the doctrine of misrepresentation. There is either a contract on one party's meaning or the other's, with any such contract possibly voidable for misrepresentation, or there may simply be no objectively corresponding offer and acceptance. *Non est factum* aside, there is no work left for a separate mistake of term doctrine to do. While *Smith v Hughes* undoubtedly recognises the category of 'known mistakes as to terms', the case itself is not a positive instance of category. The actual decision was only to order a new trial to see whether the jury's conclusion to void the contract could be defended in the light of a more precise statement of the qualifying conditions. The judges clearly did not think so.[85] On retrial, the buyer could only win if further evidence emerged to show that there was actually a contract for 'old oats', an actionable misrepresentation as to the age of the oats, or no objective agreement at all.

Two examples can be given where there is questionable or no objective coincidence of intentions necessary for contract formation. In *Centrovincial Estates plc v Merchant Investors Assurance Co Ltd*, land had been rented at £68,320 per annum.[86] When the rent was reviewed, the landlord proposed £65,000 when he meant to say £126,000; the tenant immediately accepted. The tenant was given leave to defend the landlord's action for a declaration that no binding contract had been concluded. The Court of Appeal held that the figure of £65,000 would stand *unless* the landlord could prove that the tenant knew or ought reasonably to have known that its offer was unintended when they purported to accept it. In *Chwee Kin*

have been wholly different; or even if he had said anything which shewed that he was not acting on his own inspection and judgment, but assumed as the foundation of the contract that the oats were old, the silence of the seller, as a means of misleading him, might have amounted to a fraudulent concealment, such as would have entitled the buyer to avoid the contract.'

[84] *Ibid*, at 610 (Hannen J): 'If by any means he knows that there was no real agreement between him and the promiser, he is not entitled to insist that the promise shall be fulfilled in a sense to which the mind of the promiser did not assent.' See also, *ibid*, at 607 (Blackburn CJ).

[85] Blackburn J (*ibid*, at 608) could 'not see much evidence to justify a finding for the defendant … if the word 'old' was not used. There may have been more evidence than is stated in the case; and the demeanour of the witnesses may have strengthened the impression produced by the evidence there was; but it does not seem a very satisfactory verdict if it proceeded on [the mistake] ground.' Hannen J (*ibid*, at 611) also found 'very little, if any, evidence to support a finding upon [the mistake ground] in favour of the [buyer].'

[86] Above n 37.

Keong v Digilandmall.com Pte Ltd, *Digilandmall.com's* (D's) employee mistakenly advertised a commercial laser printer for $66 on D's website (less than 2 per cent of the actual retail price of $3,854).[87] By the time the error was detected, 4,086 orders had been received and confirmation notes automatically dispatched within a few minutes. D resisted Chwee's (C's) action to enforce their order for 1,606 printers, alleging that its unilateral mistake was known to C. The Singapore Court of Appeal upheld the finding that the buyers were 'fully conscious that an unfortunate and egregious mistake had indeed been made by the defendant'.[88] Although *actual* knowledge of the mistake was said to be required, a very generous view was taken of its scope. According to the court, it can be inferred from circumstantial evidence: 'Phrases such as "must have known" or "could not reasonably have supposed" are really evidential factors or reasoning processes used by the court in finding that the non-mistaken party did, in fact, know of the error made by the mistaken party.'[89] Moreover, it includes '"Nelsonian knowledge," namely, wilful blindness or shutting one's eyes to the obvious'.[90] In both these cases, the issue was whether the parties reached agreement and what the agreement was. Talk of mistake of terms known to the other party merely imports unnecessary distraction.

The equitable analogue is *Webster v Cecil*.[91] In that case, Cecil (C) offered to sell land for £1,250 when he intended to say £2,250. The mistake must have been obvious to Webster (W) because C had previously refused to sell for £2,000 and C informed W of his mistake immediately after W's purported acceptance. The court of equity refused W's claim for specific performance, but, in the absence of sufficient evidence to hold the buyer to a contract at £2,250, common law would have found no concluded contract at all.

B. What was the Objective Agreement For?

In contrast to cases where objective agreement is absent, where D knows of C's objective meaning and leads C to believe that D is consenting to it, there is a contract on C's meaning although this deviates from the literal or ordinary

[87] [2005] SGCA 2, affirming [2004] 2 SLR 594 (HC) [*Digilandmall.com*].

[88] According to the High Court, the relevant factors were: (i) the 'stark gaping difference between the price posting and the market price of the printers'; (ii) the fact that the buyers were 'well-educated professionals—articulate, entrepreneurial and, quite bluntly, streetwise and savvy individuals'; and (iii) the fact that the printers were purchased in the 'dead of night' with 'indecent haste' and the e-mails between the purchasers showed that they were anxious to place orders before the mistake was corrected: see *Digilandmall.com* (HC), *ibid*, at [142]–[145].

[89] *Digilandmall.com* (CA), above n 87, at [35].

[90] *Ibid*, at [42].

[91] (1861) 30 Beav 62, 54 ER 812.

meaning of C's conduct. In *Hartog v Colin & Shields* the seller offered to sell 3,000 Argentine hare skins at a 10d 'per pound' when he really meant and had previously offered 10d '*per piece*'.[92] There were three pieces to the pound. The buyer purported to accept and sued for damages when the seller refused to deliver for one third of his intended price. The court's conclusion in favour of the seller has been explained in terms of the seller's subjective intention trumping his objective intention when the buyer knows of the seller's mistake as to terms (that is knows of the seller's subjective intention): he is prevented from snatching a bargain known not to have been intended for him. Again, this rests on the erroneous detached view of objectivity.[93] On the correct observer-contextual version of objectivity, an honest and reasonable buyer would interpret the seller as meaning 'per piece'. The court found that the buyer 'could not reasonably have supposed' the seller intended to quote the price 'per pound' given the practice in the trade and the pre-contractual negotiations (verbal and written), which *always* discussed the price 'per piece' and never 'per pound'.[94] Moreover, Singleton J said that he found it 'difficult to believe that anyone could receive an offer for such a large quantity of Argentine hares at a price so low as 3d per piece without having the gravest doubts of it … the plaintiff must have realised, and did in fact know, that a mistake had occurred'.[95] If the observer knows or has reason to know that the actor's meaning does not coincide with the detached objective interpretation of his or her conduct (that is has made a mistake in stating the terms of his or her offer) then, as a reasonable and just person, the observer would not treat the other party as having agreed to it. An unreasonable and unjust observer is not allowed to assert their unreasonableness and 'snap up' the offer. Indeed, in *Hartog v Shields*, it is clear that the buyer knew that the seller meant to offer 'per piece', an offer which he purported to accept. Against *that* point of reference, the seller made no mistake. In principle, he could have, but chose not to, enforce the contract on a 'per piece' basis.[96] The decision was simply that he was able to resist the buyer's enforcement of a contract on the 'per pound' basis since there was no such contract.

[92] [1939] 3 All ER 566 (KBD) [*Hartog v Shields*].

[93] See, eg, McKendrick, *Contract Law: Text, Cases and Materials*, above n 3, at 40: 'objectively, the parties to reach [an] agreement. The defendants offered to sell their hareskins at a price per pound and the plaintiff accepted that offer. It was the fact that the plaintiff knew that the defendants were mistaken that led Singleton J to conclude that no contract had been concluded to sell their hareskins at a price per pound. This suggests that the vital factor in persuading Singleton J to conclude that the plaintiff was not entitled to recover damages was his finding that the parties were not subjectively agreed.'

[94] *Hartog v Shields*, above n 92, at 568.

[95] *Ibid*, at 567–8.

[96] This is consistent with the requirements of rectification for unilateral mistake as to the wording of a document. This analysis is found in text accompanying nn 68–71. See *Bates v Wyndhams*, above n 68, at 515–16, 520–21; *New Towns v Cooper*, above n 67.

Scriven v Hindley can be interpreted in the same way.[97] It also instances the objective approach's bias to the just and reasonable. The buyer successfully bid, but refused to pay, for an auction lot he believed to contain hemp when it actually contained the much cheaper commodity tow. Lawrence J found no contract on the unnecessarily wide basis that 'the parties were never *ad idem* as to the subject-matter of the proposed sale', one intending to sell tow and the other intending to buy hemp. However, here, the *non*-coincidence of the parties' 'subjective' meanings, which was unknown to either party, would have been irrelevant[98] without the seller's misleading auction catalogue. This described the goods as so many bales in different lots, all bearing the same shipping marks which, witnesses explained, never happened before for different commodities from the same ship. The bidder, quite reasonably, did not to foresee the potential for confusion: being only interested in buying hemp, he had only inspected the hemp on show and not the tow bearing the same shipping marks. Lawrence J held that since the confusion was deliberately perpetrated by the seller-observer to swindle the bank financing the shipment, 'it was peculiarly the duty of the auctioneer to make it clear to the bidder . . . which lots were hemp and which lots were tow'.[99] While auctioneers are generally entitled to assume that bidders know what they are bidding for (their mistakes are legally irrelevant), they cannot do so if they have carelessly, albeit unintentionally, induced the bidder's mistake. The case is authority for observer-contextual objectivity, *not* for the priority of subjective intention over detached-formal objectivity. It shows that where the *seller's* conduct gives the bidder reason to believe that he or she is bidding for hemp the seller cannot treat the bidder's offer as for tow. Indeed, there is arguably a contract for hemp.

The equitable analogue is *Denny v Hancock*.[100] In that case, the purchaser inspected the property with the assistance of plans that showed one side to be bounded by trees. He naturally concluded that three magnificent trees going up to an iron fence were inside the property; in fact they were not, the real boundary being denoted by stumps made inconspicuous by shrubs. In denying specific performance, James LJ stressed that any reasonable prospective purchaser would be misled. He said: '[i]f I had done exactly what this gentleman did, and taken their plan in my hand, and gone through the property … I should have arrived at exactly the same conclusion as this gentleman did.' Although the focus was on whether the purchaser could resist specific performance, it is arguable that if there was any objective agreement, it was for the property as the seller knew that the

[97] [1913] 3 KB 564, 568.
[98] *Robinson, Fisher and Harding v Behar* [1927] 1 KB 513.
[99] *Scriven v Hindley*, above n 97, at 569.
[100] (1870) LR 6 Ch App 1, 11.

buyer would have understood and intended to contract for. At common law, the buyer should have been entitled to compensation for his loss of expectation.

C. Latent Ambiguity

In the well-known case of *Raffles v Wichelhaus* the parties contracted to buy and sell goods 'to arrive ex Peerless from Bombay', the buyer intending the ship *Peerless* arriving in October while the seller delivered on another ship, also called *Peerless*, arriving in December.[101] The court upheld the buyer's refusal to pay but without giving its reasons. The case is interpreted as an exception to the objective test, either on the basis that the parties' *subjective* intentions do not coincide[102] or that there is a mutual mistake in that each party is mistaken as to the other's intention.[103] With respect, both explanations are unnecessarily wide.

Exact subjective coincidence of intentions must be rare in contracting; non-correspondence is not generally a ground for voiding the contract. The objective test, *by definition*, makes irrelevant any party's subjective mistake about the other party's intention. What matters is each party's *evident* intention honestly, reasonably and contextually interpreted from the observer's perspective. On this approach, offer and acceptance corresponded; there *was* agreement to buy and sell goods 'to arrive ex Peerless from Bombay'. The problem was identifying which of the two ships fitting that description was the subject of the agreement, which version should be enforced. The answer was 'don't know, can't say'. It belongs with other cases where vagueness (rather than lack of agreement) prevented contract formation, as where agreements to sell goods 'on hire-purchase terms'[104] or 'subject to war clause'[105] but the courts could not say which of the many different, and each reasonable, versions of such terms the parties intended. The parties *were* objectively agreed, but the agreement suffered from latent ambiguity which was impossible to resolve by reference to the context: objectivity simply 'ran out'. Even the more expansive contextual objectivity is not omniscient.

[101] *Raffles v Wichelhaus*, above n 37.
[102] See Beatson, above n 3, at 321–2 who discusses the case under the heading of 'Absence of genuine agreement' and 'Offer and acceptance not coincident'.
[103] See Furmston, above n 32, at 306.
[104] *Scammell and Nephew Ltd v Ouston* [1941] AC 251 (HL).
[105] *Bishop & Baxter v Anglo-Eastern Trading and Industrial Co Ltd* [1944] KB 12 (CA).

D. Actor Objectivity: *Non est factum*

The *non est factum* doctrine comes closest to appearing to take account of contract parties' subjective intentions. Where one party's mistake as to the nature of a signed contractual document is 'fundamental', or 'essential', or 'radical', or 'very substantial', or 'serious,'[106] *non est factum* can void the contract even if the mistake is unknown to the other party (if it was, there would be no objective agreement as in *Smith v Hughes*) and has not been induced by the other party's misrepresentation (if it was, the contract would be voidable even if the mistake was not 'fundamental').[107] But *non est factum* is not an instance of the exceptional priority of subjectivity; the doctrine represents an exceptional switch from the perspective of the observer to that of the *actor*, although the actor in question must suffer the relevant cognitive disability[108] and not have been careless.[109] The test is still *objective* since the claimant must show that his mistaken belief about the nature of the document was honest and reasonable for someone with that disability in the circumstances.

The doctrine is justified on the basis of no consent,[110] but this is over-inclusive. Specifically, what justifies the switch from the observer to the actor's perspective? The answer is that while contract law aims to facilitate the autonomy-enhancing institution of contract by preventing its abuse (hence *observer*-objectivity which safeguards voluntariness at the *institutional* level),[111] contract law cares about autonomy at the *individual* level and hence *actor*-objectivity, to this extent. The actor's conduct which

[106] *Saunders v Anglia Building Society* [1971] 1 AC 1004 (HL) [*Saunders*], affirming *Gallie v Lee* [1969] 2 WLR 901 (CA).

[107] See *United Dominions Trust Ltd v Western* [1976] QB 513 (CA) [*United Dominions Trust*]. In *United Dominions Trust*, the defendant's hire purchase agreement, tainted by the dealer's fraud, would have been void against the finance company although the latter was ignorant of the fraud, had the defendant not been careless in signing a blank agreement and had the agreement been fundamentally different from that intended by the defendant.

[108] *Saunders*, above n 106, at 1016: be 'permanently or temporarily unable through no fault of their own to have without explanation any real understanding of the purport of a particular document, whether that be from defective education, illness or innate incapacity,' or from being tricked.

[109] In *Saunders*, *ibid*, a widow who had broken her glasses was barred by her carelessness in not checking the identity of the transferee. The requirement of care would disqualify a claimant who signs a document in blank leaving another to fill in the details, see *United Dominions Trust*, above n 107.

[110] See *Foster v Mackinnon* (1869) LR 4 CP 704. At 711, Byles J said the contract is invalid 'on the ground that the mind of the signer did not accompany the signature: in other words, that he never intended to sign, and therefore in contemplation of law never did sign, the contract to which his name is appended'. See also *Hasham v Zenab* [1960] AC 316 (PC) 335; *Mercantile Credit Co Ltd v Hamblin* [1965] 2 QB 242 (CA) 268, 280. *Saunders*, above n 106, at 1026.

[111] See text accompanying nn 14–16.

counts as consenting to a contract must be: (a) intentional;[112] (b) accompanied by knowledge that the conduct will count as contracting;[113] and (c) accompanied by a minimum threshold of accuracy as to the gist of the contract.

In relation to the first criteria, an action will count as intentional if the actor knowingly engages in the conduct. Thus, in this area the law imposes a very weak form of intent which covers almost everything short of automatism, sleep walking, being pushed and so on.

In relation to the second criteria, knowledge that the conduct will count as contracting is absent if, for example: (a) the actor believes he or she is giving someone an autograph when the actor is actually signing a contract; or (b) the actor drives into a car park believing it to be free when it is charged, even if the other party reasonably believes the actor is consenting to a contract.

In relation to the third criteria, effective consent does not require accuracy on every *detail* of the content (hence, the observer-objective test of intention, the signature rule, the rules on incorporation and the curing of uncertainty in formation and so on). Nevertheless, if a claimant can show that he or she has made a fundamental mistake as to the very *nature* of the obligation undertaken, then the claimant can claim that the consent was not meaningful enough to bind him or her. The same reasoning justifies voiding a contract for common fundamental mistake as to fact (assumption) since this radically defeats the *purpose* or the *means* for achieving the purpose of the contract.[114] Of course, a very high threshold of seriousness is required (the difficulty of defining and applying this is notorious). Other hurdles may be imposed to protect the certainty and security required of a valuable institution of contract. But the idea is that a contract law which enforces voluntarily assumed obligations should, at least, accept that a party who is so fundamentally mistaken about the *gist*, or *core*, or *substance* of what he or she has undertaken (he or she is not even in the right 'ball park') has not really undertaken anything. Phrased differently, you need not actually agree to every rule of the club to be bound by them, but you must at least have joined the right club.[115]

[112] Goddard, above n 16; Sheinman, above n 22.
[113] Endicott, above n 16; Sheinman, above n 22.
[114] See text accompanying nn 156–8.
[115] See further De Moor, above n 44, at 635–55; Endicott, above n 16, at 151; Raz, above n 13, at 933; Sheinman, above n 22.

E. Mistaken Identity

What of the 'notoriously unsatisfactory'[116] English law on mistake as to the other party's identity? The central problem is the rogue posing as someone else who buys goods with a worthless cheque or on credit, sells the goods on and then absconds before his or her fraud is discovered. Should the original owner be able to recover the goods from an innocent third-party purchaser? The plethora of potentially applicable rules, the usual presence of misrepresentation and signatures and the concern to protect innocent third-party purchasers have yielded a set of potentially contradictory and ossifying guiding rules. Thus, the contract is said to be: (a) void if the mistake is as to the other party's *identity* but voidable if the mistake is only as to that party's' *attribute*; and (b) void if the contract is made 'in writing',[117] but voidable if it is made 'face-to-face'.[118] These rules of thumb wrongly suggest that the contract is void if the claimant is very mistaken (the claimant wins), but voidable if he or she is only a little mistaken (the third party wins).

The proper starting point must be the basic rule of contract formation that no one can accept an offer which he or she knows or has reason to know is not intended for that party. Neither can a party rely on an apparent acceptance knowing it was in response to an offer believed to have come from someone else.[119] If the other party makes an offer or accepts an offer under mistake as to the offeror's identity, the question is one of formation—whether a contract has been made at all—not whether the mistake should vitiate any prima facie valid contract. This assumes that the identity of the other contract party is important to the claimant.[120] If it is not, (for example, retailers are not generally concerned about the identity of the shopper, or the auctioneer with the identity of the bidder),[121] the contract comes into existence despite the mistaken identity although it may be voidable if it was nevertheless induced by the rogue's misrepresentation.

[116] *Shogun Finance Ltd v Hudson* [2004] 1 AC 919 (HL) [1], [34] (Lord Nicholls). For a similar view, see Law Reform Committee, *12th Report*: *Transfer of Title to Chattels* (London, Her Majesty's Stationary Office, 1967); G McMeel, 'Interpretation and Mistake in Contract Law: "The Fox Knows Many Things…"' [2006] *Lloyd's Maritime and Commercial Law Quarterly* 49; C MacMillan, 'Rogues, Swindlers and Cheats: The Development of Mistake of Identity in English Contract Law' [2005] *CLJ* 711; C MacMillan, 'Mistake as to Identity Clarified?' (2004) 120 *LQR* 369; C Hare, 'Identity Mistakes: A Missed Opportunity?' (2004) 67 *MLR* 999; AL Diamond, 'Law Reform Committee: Twelfth Report on the Transfer of Title to Chattels' (1966) 29 *MLR* 413.

[117] *Cundy v Lindsay* (1878) 3 App Cas 459 (HL).

[118] *Phillips v Brooks* [1919] 2 KB 243.

[119] *Boulton v Jones* (1857) 2 H & N 564, 157 ER 232 (Ex Ct).

[120] *Ingram v Little* [1961] 1 QB 31 (CA) 57; *Lewis v Averay* [1972] 1 QB 198 (CA) 209.

[121] *Dennant v Skinner* [1948] 2 KB 164.

However, even if vital to the claimant, the identity of the other party to the contract does not quite fit the description of 'term'; it is artificial to say that the rogue promises his identity as a term of the contract. Nevertheless, while the identity of one's contract partner does not go to *what* the contract is for, it does go to the equally important question of *who* one chooses to contract with (to whom one has chosen to undertake contractual obligations). In this sense it is very much part of the contract formation question. We can regard it as a condition precedent, a term of the claimant's offer, or of his or her acceptance. On this view, and consistently with observer-contextual objectivity, claimants should be able to deny contracts if they can show that: (i) they were mistaken about the other party's identity; (ii) that identity was vital to them;[122] and (iii) the other party (the rogue) knew or had reason to know of (i) and (ii).

Proof of (ii) would be analogous to that required to show that a statement made during negotiations is a term or collateral term of the contract, rather than a mere representation.[123] Identity is clearly vital where, for example, an offer is made only to persons fitting particular descriptions which exclude the rogue ('current students of a particular university', or being 'over 18 years of age'); or, where the rogue knows from previous dealing that the claimant is unwilling to contract with him or her (since he or she is barred from a pub or a soccer match).[124] The claimant must show that 'but for' the mistake the claimant would not have entered the contract. In contrast, a misrepresentation is operative if it was merely *a* reason for the claimant entering the contract. On Lord Millett's scenario in *Shogun Finance Ltd v Hudson*, the contract is voidable where a man books a hotel room 'for himself and a girlfriend under a common but fictitious name in order to give the impression (when such things mattered) that they were married'.[125] But the contract may be void if the man knows that the hotel would only accept married couples. Logically, mistake about any attribute of the rogue which is vital to the claimant should count. However, his creditworthiness should be excluded because if the claimant allows payment by cheque or extends credit, this is simply a business risk that it takes.

As for (iii), where the rogue has misrepresented his or her own identity, this is weighty evidence that the rogue not only knows of the claimant's mistake (which he or she has induced), but also that the rogue knows of the importance of his or her identity to the claimant (hence the incentive to

[122] *Ingram v Little*, above n 120, at 57; *Lewis v Averay*, above n 120, at 209.

[123] *Heilbut Symons & Co v Buckleton* [1913] AC 30 (HL) [*Heilbut Symons*]; *Oscar Chess Ltd v Williams* [1957] 1 WLR 370 (CA); *City of Westminster Properties (1934) Ltd v Mudd* [1959] Ch 129 [*Westminster Properties v Mudd*], and see text accompanying nn 145–7.

[124] *Said v Butt* [1920] 3 KB 497.

[125] Above n 116, at [78].

lie). *Boulton v Jones* shows that the rogue may have the relevant knowledge even without an active misrepresentation.[126] The determination should be made without undue distortion by the presumptions that contractual intention is present when a rogue's misrepresentation occurs in the claimant's presence (face-to-face), but absent when it is made in the claimant's physical absence (and the contract is constructed from an exchange of writing or is reduced to a written document). If a party was ignorant that the claimant had mistaken his or her identity or of the importance of this identity to the claimant then the contract comes into existence (for example, I enter a contract with you only because I mistakenly believe you are a famous actor). A party may be similarly ignorant even if he or she has misrepresented this identity. For example, if, in the hotel scenario, the man can show that his lie is not primarily aimed at deceiving the hotel but to mislead, say, journalists or his wife. That leaves voidness due to common (shared) fundamental mistaken assumption, but this requires a factual scenario which is impossible to imagine (since the rogue knows his or her own identity) without some medical condition akin to amnesia.

On this approach, mistaken identity cases should be resolved via contract formation analysis. If this approach is taken, many more cases would result in a conclusion that there was no contract. The concern with third party rights, a concern which rightly belongs in property law, should be dealt with separately and not be allowed to twist contract law principles which are aimed at assessing the rights between the contracting parties. The tail should not wag the dog.

III. DISTINGUISHING MISTAKEN ASSUMPTIONS

A. Imposing Taxonomic Order: Formation and Vitiation

Smith v Hughes leaves us with a final conundrum: why is it that the buyer's mistaken belief that the oats he purchased were old attracts no relief (because it is not shared by the seller nor sufficiently fundamental in importance), but the buyer's mistake that the seller had *promised* the oats to be old would void the contract if known to the seller? If a mistake as to some non-fundamental quality of the subject matter does not negate consent, why should a mistake as to whether the other party promised it do so? Indeed, one might go on to ask how we should understand the

[126] Above n 119. In that case, J sent a written order for some goods to Brocklehurst, who J had dealt with previously and against whom he could set off sums that Brocklehurst owed him. B took over Brocklehurst's business and filled J's order without disclosing the change of ownership.

different requirements of the formidable list of different mistakes commonly identified in case law and commentary.[127]

The problem is one of unstable classification. The categories of mistake cut across each other; some are about *who* has made the mistake, some about the *type* of mistakes made, some about the *seriousness* of the mistake, and some about the different *requirements for relief* under the common law and at equity (for example, knowledge, fault on the defendant's part or lack of fault on the claimant's part). The same scenario may be susceptible to different descriptions to which different legal consequences attach. Absence of a dominant taxonomy in the law of contractual mistake impedes understanding, leaving the student or judge to take refuge in a list approach.

Clarity about the underlying structure of a subject is vital to clear thinking, principled development of the law and the eradication of inconsistencies. As Lord Steyn reminds us, 'in law classification is important. Asking the right questions in the right order reduces the risk of wrong decisions.'[128] A model of mind-boggling taxonomy which conforms to no single classifying scheme is the categorisation of animals identified by Jorge Luis Borges in 'a certain Chinese encyclopaedia'. There:

> animals are divided into (a) belonging to the emperor, (b) embalmed, (c) tame, (d) suckling pigs, (e) sirens, (f) fabulous, (g) stray dogs, (h) included in the present classification, (i), frenzied, (j) innumerable, (k) drawn with a very fine camelhair brush, (l) et cetera, (m) having just broken the water pitcher, (n) that from a long way off look like flies.[129]

What classificatory rationale underpins different types of mistakes and so explains the different requirements for relief? To unravel the tangle we need to start with another capacious idea in play here; namely, that when contracts are set aside, including for mistake, it is because of the claimant's 'defective consent' to the contract. Like mistake, 'defective consent' is not precise enough to be useful, but is loose enough to cause real trouble. HLA Hart warned that the process of reasoning which holds that since consent gets you *into* contract, only lack of consent will get you *out* is 'a disastrous

[127] A non-exhaustive list of mistakes includes: common mistake, unilateral mistake, mutual mistake, cross purpose mistake, mistakes going to the root of the contract, absence of genuine agreement, common law mistake, equitable mistake, mistake as to the identity or intention of the other party, fundamental mistake as to the identity, ownership, existence, quality or quantity of the subject matter, fundamental mistaken assumption, mistaken recording of the contract (rectification), and fundamental mistake about the nature of the documents signed (*non est factum*).

[128] *Attorney-General v Blake* [2001] 1 AC 268 (HL) 290.

[129] M Foucault, *The Order of Things* (New York, Pantheon Books, 1970) xv, referred to in D Johnston and R Zimmermann, 'Unjustified Enrichment: Surveying the Landscape' in D Johnston and R Zimmermann (eds), *Unjustified Enrichment: Key Issues in Comparative Perspective* (Cambridge, Cambridge University Press, 2002) 3, 25–6.

over-simplification and distortion' of the law governing the vitiation of transactions.[130] This reasoning fails to recognise that consent is a necessary but not sufficient, condition of contractual liability, and that the enforceability of a contract is a two-stage inquiry. Even when the language of consent is used in both the formation and vitiation stages, they deal with qualitatively different concerns. 'Formation' asks whether the parties have succeeded in reaching agreement. Here, the commitment and content questions are merged in the offer and acceptance approach. In contrast, 'vitiation' asks whether, in view of the relevant circumstances, a claimant should be released from liability under the contract, despite her consent to it.[131]

As Hart explains:

> When the student has learnt that in English law there are positive conditions required for the existence of a valid contract … his understanding of the legal concept of a contract is still incomplete … For these conditions, although necessary, are not always sufficient and he has still to learn what can *defeat* a claim that there is a valid contract, even though all these conditions are satisfied. That is the student still has to learn what can follow on the word 'unless' which should accompany the statement of these conditions … the law has a word which with some hesitation I borrow and extend: This is the word '*defeasible*' used of a legal interest in property which is subject to termination or '*defeat*' in a number of different contingencies but remains intact if no such contingencies mature. In this sense then, contract is a defeasible concept.[132]

The truth of these statements is obscured by those who take the language of impaired consent as suggesting 'that there are certain psychological elements required by the law as necessary conditions of contract and that the defences [vitiating factors] are merely admitted as negative *evidence* of these', rather than as 'a compendious reference to the defences with which claims in contract may be weakened or met'.[133] In short, contract formation is not conditional on the parties' minds being free from mistake; the party seeking to enforce a contract does not have the onus of proving the absence of mistake on the other's part. The claimant's success in voiding a contract for mistake is not proof of lack of consent to contract formation.

[130] HLA Hart, 'The Ascription of Responsibility and Rights' (1948) 49 *Proceedings of the Aristotelian Society* 171, 183 (reprinted in AGN Flew (ed), *Logic and Language—First Series* (Oxford, Blackwell Publishing, 1963)).

[131] *Ibid*, at 174.

[132] *Ibid*, at 174–5 (emphasis in the original).

[133] *Ibid*, at 177. At 180, Hart explains that 'the logical character of words like "voluntary" are anomalous and ill-understood. They are treated in such definitions as words having positive force, yet, as can be seen from Aristotle's discussion in *Book III of the Nicomachean Ethics*, the word "voluntary" in fact serves to exclude a heterogeneous range of cases such as physical compulsion, coercion by threats, accidents, misstakes, etc, and not to designate a mental element or state; nor does "involuntary" signify the absence of this mental element or state.'

In these cases, the contract is set aside despite consent because the presence of, for example, duress[134] or undue influence[135] means that the law does not ascribe the normal responsibility it would to the victim's consent. His or her consent is only *deemed* to be defective. Talk of defective consent in the context of vitiating factors is conclusory, not explanatory; it is, as Hart reminds us, merely a shorthand for the variety of factors rendering a transaction defeasible.

B. Terms and Assumptions

The primary taxonomic distinction between (a) contract formation and (b) vitiation of contract is fundamental. It mirrors that key distinction between *terms* inside a contract and *assumptions* (non-terms, or 'mere representations') outside the contract. The distinction can be exceedingly difficult to draw.[136] As *Smith v Hughes* shows, the quality of the subject matter (that the oats are old) may be either a term or an assumption, attracting very different tests for relief. The same applies to the existence,[137] or identity[138] of the subject matter of the contract.

The distinction between terms and facts is vital in determining whether false statements in pre-contractual negotiations attract the consequences of breach of contract or for misrepresentations inducing agreement to the contract. Since misrepresentations are but induced mistakes, it stands to reason that the distinction should be equally important in the law of mistake.

(i) Mistake of Terms and Mistakes of Fact

Mistake of terms and mistakes of fact raise entirely different issues and invoke different principles. So-called 'mistake as to terms' goes to the *formation* and *contents* of the contract. These cases raise the issues of offer

[134] This resonates with the rejection of the 'overborne will' explanation of duress in *Universe Tankships Inc of Monrovia v International Transport Workers Federation* [1983] 1 AC 366 (HL). As Lord Scarman said at 400: 'The classic case of duress is … not the lack of will to submit but the victim's intentional submission arising from the realisation that there is no other practical choice open to him.'

[135] See *Royal Bank of Scotland v Etridge (No 2)* [2002] 2 AC 773 (HL) [7] where Lord Nicholls said that whenever the defendant's procurement of the claimant's consent is judged improper by the standards of the undue influence doctrine, that consent will not be *deemed* an expression of the claimant's will.

[136] See *Heilbut Symons*, above n 123; *Oscar Chess Ltd v Williams*, above n 123.

[137] It was classed as a term in *McRae v Commonwealth Disposals Commission* [1951] 84 CLR 377 (HCA) and has been interpreted as a fact in *Couturier v Hastie* (1856) 5 HL Cas 673.

[138] It was classed as a term in *Scriven v Hindley*, above n 97, at 567 and as a fact in *Leaf v International Galleries* [1950] 2 KB 86 (CA).

and acceptance (objectively determined), implied terms and collateral terms. Moreover, breaches of contractual terms trigger forward-looking remedies in the nature of vindicating expectations.

Mistakes of fact are mistaken assumptions about the *context* in which the contract is made. These mistakes affect a party's motivation for, or assessment of, the desirability of the contract. 'Mistakes of fact' go to the *vitiation* of contract. It asks whether—despite the parties' objective consent to the contract—one party should be excused from contractual liability. The remedy is backward looking and, broadly speaking, aimed at restoring the parties to their pre-contractual position.

(ii) Mistake and Objectivity

The word 'mistake' connotes deviation from the accurate (unmistaken) point of reference. Talk of 'mistake' makes sense when applied to assumptions, less so to contractual terms. A party makes a mistaken assumption if his or her evident (and in that sense 'objective') belief about the matter deviates from the *independently verifiable truth* of the matter. You believe the oats you physically inspected and bought are old but they are not. I believe that the sea vessel I hired to standby and evacuate the crew of my distressed ship pending the arrival of a rescue tug is 35 miles away; in fact it is 400 miles away.[139] You lease a room to watch the Royal Coronation procession, but it has already been cancelled.[140] I pay to terminate your contract of employment not realising that I could have done so without paying.[141]

In contrast, so-called 'mistake of terms' are not so straightforward. Unlike facts, the reference point of contractual terms has no independently verifiable existence. It is the *construct of the parties* and emerges from the coincidence of each party's objective interpretation of the other's intention. In *Smith v Hughes* the evidence pointed to an agreement for 'oats'. Since the seller had reason to believe that that was the buyer's intention, he could not simultaneously know (that is have the inconsistent belief) that the buyer really intended to buy 'old oats'. The buyer's belief is treated as one of factual assumption, which only voids the contract if it is fundamental and shared by the seller.

Even where a contractual document purporting to contain the parties' agreement is present, the question is whether each party has given the other reason to believe that he or she intends to be bound by the four corners of the document, hence the exceptions to the signature rule and the parole evidence rule and the remedy of rectification. If not, the court may be able

[139] *Great Peace Shipping v Tsavliris* [2003] QB 679 (CA) [*The Great Peace*].
[140] *Griffith v Brymer* (1903) 19 TLR 434 (KBD).
[141] *Bell v Lever Brothers Ltd* [1932] AC 161 (HL).

to add implied or collateral terms that may even override the terms of the contractual document;[142] or may rectify the contractual document to bring it into line with the parties' evident and agreed intention.[143] In each case, the parties' *real* agreement, if provable, trumps the different written record of their agreement. Thus, the contractual document will only be the objective reference point for determining mistake if, objectively determined, the parties have agreed that it embodies their agreement *and* their agreement has not been wrongly recorded. The resulting potential for circularity should not obscure the fact that in ascertaining whether there has been a mistake of terms and who has made it, it is not necessarily enough to just look for deviation from any contractual document. The point of reference is the parties' corresponding objective intentions judged from the observer's perspective, in the context of their dealings and assuming honesty and reasonableness on all sides.

(iii) Contract 'Void' for Mistake

The importance of the formation/vitiation and term/fact distinctions is obscured by the description that contracts tainted by mistake are 'void' at common law. This word fails to distinguish between two categorically different situations. The first situation is one where no contract ever comes into being because there was no corresponding offer and acceptance from the objective point of view.[144] In the second situation, there is a prima facie valid contract that is not binding in the circumstances since it is defeasible because vitiated for mistake of fact. The distinction between mistakes which 'negative' consent and those which 'nullify' consent, made by Lord Aitkin's judgment in *Bell v Lever Brothers Ltd*, reaches towards the same distinction but it is not sufficiently transparent to be helpful.[145]

C. Fundamental Common Mistaken Assumption

Since 'mistake of terms' is about contract formation, it is entirely logical that *any* objectively determined non-correspondence of the parties' offer and acceptance should prevent contract formation (no contract results from my offering to sell *new* oats and you purporting to accept *old* oats).

[142] *Westminster Properties v Mudd*, above n 123, *Mendelssohn v Normand Ltd* [1970] 1 QB 177 (CA).

[143] *Lovell and Christmas*, above n 65. Even the presence on an 'entire agreement' clause in the contract does not prevent rectification, see *JJ Huber*, above n 65.

[144] There may also be a contract but not on the terms sought to be enforced (because one party knew or had reason to know that the other's intention did not correspond with the words used *and* what the latter intended). See text accompanying nn 84–91.

[145] Above n 141, at 217, cited, eg, in Peel, above n 4, at [8–001].

An attempted agreement has failed. On the other hand, contract parties must generally take the risk of mistaken assumptions outside that guaranteed by the contract terms. This is necessary to protect the security of transactions and contract parties' reasonable expectations (that is the institution of contract). Hence, the contract stands if it is for new oats (or just 'oats') but you merely believe (assume) they are old.

The correct classification shows that the real question is *not* why a fundamentality requirement is absent for operative mistake of terms,[146] but why relief should be given for mistakes of fact at all. The law shows that the legitimate concern to protect the security of contracts and the parties' reasonable expectations is outweighed when two conditions are satisfied. First, a claimant must show that his or her mistake was sufficiently serious ('fundamental') to divest their consent of significance in the actual (unmistaken) circumstances. The mistake may, for example, make the contract pointless[147] or unachievable.[148] A claimant's cry is not really that he or she did not consent, rather, it is that he or she *did* consent, but not to performance in *these* circumstances, the risks of which the claimant neither anticipated nor expressly or impliedly assumed. The analogous doctrine of frustration discharges a contract *from* the occurrence of the frustrating event because that is when a claimant's consent 'runs out'.

Second, the claimant must show that the other party shared this mistake. Pragmatically, this corroborates the claimant's assertion of mistake (all too easy to make) and the importance of the mistake.[149] More importantly, this removes the other party's claim to the protection of his or her expectation since that is *also* tainted by the catastrophic mistake. Voiding the contract in such circumstances deprives the other party of benefits (including unexpected windfalls) which are *unworthy of protection* because the other

[146] Peel, above n 4, at [8–044].

[147] For example, buying one's own property as in *Cooper v Phibbs* (1867) LR 2 HL 149.

[148] See, eg, *The Great Peace*, above n 139. In that case, the question was whether the GP hired by T was actually so far away from T's distressed ship 'at the time of the contract as to defeat the contractual purpose [to provide escort and standby services for five days until the rescue tug arrived]—or in other words to turn it into something essentially different from that for which the parties had bargained? This is a question of fact and degree': *Great Peace Shipping Ltd v Tsavliris Salvage (International) Ltd* (2001) 151 NLJ 1696 (QBD) [56] (Toulson J)). The Court of Appeal concluded that although the GP was 400 miles (39 hours sailing) away, rather than 35 miles (three hours sailing) away as believed, it was not so far away as to defeat the parties' common assumption that it could still render the rescue service desired. This was reinforced by the T's failure to cancel the agreement on discovering the true distance between the ships until they found and hired a nearer vessel to assist. On the other hand, Toulson J said that if there was five day's sailing distance between the ships, the contract *would be void* since its purpose would be unachievable: *ibid*, at [55].

[149] Shared assumptions are more likely to relate to the *essential substance* of the contract, whilst unilateral mistakes are more likely relate to matters of importance only to the claimant (and not going to the substance of the contract).

party could not reasonably have expected them when he or she entered the contract. The same analysis applies to contracts discharged for frustration.[150]

D. Misrepresentation

What difference, if any, does it make if one party's alleged 'mistake' is induced by the other party's statement? Recognition that misrepresentations merely induce mistakes should be reflected in any coherent legal scheme. *If* the buyer's 'mistake' in *Smith v Hughes* had been as to a term (that is he thought the seller was promising 'old oats'), and *if* the mistake had been induced by the seller's statement, then the buyer should be entitled to treat the seller as having consented to sell old oats.[151] The buyer can resist the seller's claim for payment, or sue for non-delivery, of the old oats. This is consistent with *Curtis v Chemical Cleaning & Dyeing Co Ltd*[152] and the remedy of rectification for common mistake in recording the parties' agreement discussed above.[153] Logically, the case for rectification should be stronger, not weaker, if one party's misrepresentation, even if innocent, induced the shared belief in the irregular meaning of words used in the document.[154]

We have noted that if the claimant enters a contract under a *unilateral* mistake as to the nature of the document or the identity of the other party known to the latter, or a *shared* mistake as to fact or assumption, the contract will only be void if the mistake was 'fundamental'. In contrast, a misrepresentation need not be fundamental to be actionable, but it only renders the contract voidable;[155] the right to rescission being vulnerable to various 'bars'. Coherence demands that the claimant should be able to *void* the contract if his or her *induced* mistake about the nature of the contractual document or the identity of the other party is fundamental. The contract should also be void if the claimant's mistaken assumption is fundamental and induced by the other party's innocent misrepresentation (so that the mistake can be characterised as common or shared). No principle of subsidiarity should apply here.

[150] See M Chen-Wishart, *Contract Law*, 2nd edn (Oxford, Oxford University Press, 2008) 290–91.

[151] See the analogous discussion in text accompanying nn 81–3, 136–8.

[152] Above n 70.

[153] See text accompanying nn 65–70.

[154] See to the contrary *Rose v Pim*, above n 31.

[155] Damages are also available for negligent, fraudulent and 'Section 2(1) of the Misrepresentation Act 1969' misrepresentations.

Misrepresentation should not trump or expel a concurrent action for mistaken identity, *non est factum*[156] or mistake in fundamental assumptions. The claimant should not be worse off because his or her fundamental mistake was induced by the defendant rather than being spontaneous. Conversely, the defendant should not be better off by inducing the claimant's mistake. In principle, this will make more contracts void when the concern to protect innocent third party purchasers pushes the other way, in the direction of voidability. The latter concern is legitimate but should be dealt with separately and transparently, rather than via twisting and manipulating contract law principles. As mentioned earlier, the tail should not wag the dog.

IV. CONCLUSION

I began with just one troublesome case. Following its thread drew me deep into the mire of the law of contractual mistake because, in a sense, *Smith v Hughes* is symptomatic of that tangled area of the law. From this exegesis, three broad conclusions emerge. The first is that we need to take great care to define what we mean by the 'objective' or 'subjective' tests of intention. Needless muddle can arise when the label is used to designate quite different points of reference. The second conclusion is that once the contours of the objective test are properly understood, the cases hitherto characterised as 'mistake of terms', including *Smith v Hughes*, are revealed as straightforward applications of objectivity. There is no need to resort to the incoherent interpretation of exceptional subjectivity trumping the objective approach. Indeed, such an explanation becomes nonsensical. The doctrine of rectification emerges as the natural corollary of this objective test of intentions. The third conclusion is that the attempt to understand why mistakes of term only have to be known by the other party (yet mistakes of fact must be shared and fundamental) in order to void contracts, requires us to stabilise the language used in the law of contractual mistake. The looseness of the key descriptive and prescriptive terms—'mistake', 'defective consent' and 'void'—allows quite different problems to be thrown together causing inevitable confusion. Locating the primary distinction between *formation* and *vitiation* allows more precise distinctions to be drawn: (a) 'defective consent' in the sense of 'no objective agreement' as opposed to 'agreement but defeasible in the circumstances'; (b) 'void' in the sense of 'no contract to start with' as opposed to 'prima

[156] The majority in the Court of Appeal in *Lloyds Bank plc v Waterhouse* (1991) 10 Tr LR 161 (CA) said that where the case is one of fraud or misrepresentation by the other party to the contract, with no third party involved, the case should be dealt with as one of misrepresentation.

facie valid contract but set aside'; and (c) 'mistake of terms' in the sense of 'looking for correspondence in each party's objective interpretation of the other's intention regarding the obligations undertaken', as opposed to 'mistake of fact' as 'mistaken assumptions not guaranteed by the contract'.

Identifying these points of reference helps to explain why known non-correspondence on *any* term prevents contract formation, while mistaken assumptions (unless induced by misrepresentation) must be shared and fundamental to void a contract. Moreover, we can begin to see how the related areas of *non est factum*, mistaken identity and misrepresentation should be classified and how troublesome issues under these headings should be resolved. The discussion can be summarised in the table below. Making sense of *Smith v Hughes* has been like untangling a giant knot. All the kinks are not straightened out but the picture looks promising.

TABLE: SUMMARY OF ANALYSIS

Formation: 'Mistake of term' (Whether and what agreement?) Point of reference: Observer-contextual objectivity except for non est factum applying actor-contextual objectivity. 'Void': no contract formation or no contract on the terms sought to be enforced.	**Vitiation: Mistaken assumptions or 'mistake of fact'** (Whether void or voidable for mistake?) Point of reference: independently verifiable fact. 'Void': prima facie valid contract vitiated.
Observer-contextual objectivity (i) Void if **'known mistake of terms'**: no objective agreement on contract terms (ii) Void if **mistaken identity**: no agreement to contract with the other party (iii) Void if ***non est factum***: fundamental unilateral mistake about the nature of document (as actor-contextual objectivity) and actor under disability and not careless (iv) Void if **incurable uncertainty** (v) **Rectification** of contractual document at equity if: —mistake unilateral and other party more blameworthy —both parties agreed to different words or meaning —misrepresentation	(i) Void if **common fundamental mistake** (ii) Voidable if **'equitable mistake'**, ie less fundamental common mistake but still serious and shared (doubtful after *The Great Peace*); or if unilateral but the other party is more blameworthy (iii) Voidable for **misrepresentation**

15

From Morgan *to* Etridge: *Tracing the (Dis)Integration of Undue Influence in the United Kingdom*

RICK BIGWOOD*

I. INTRODUCTION

'THE LAW IN general leaves every man at liberty to make such bargains as he pleases,' wrote Sir John Salmond in *Brusewitz v Brown*,[1] 'and to dispose of his own property as he chooses.'[2] He continued:

> However improvident, unreasonable, or unjust such bargains or dispositions may be, they are binding on every party to them unless he can prove affirmatively the existence of one of the recognized invalidating circumstances, such as fraud or undue influence.[3]

Granted, undue influence has long been recognised as a substantial reason for setting aside an objectively concluded transaction, be it contract or gift,[4] but what, when he wrote these words in 1922, did Salmond J understand 'undue influence' to mean?

* I am grateful to Joachim Dietrich, Graham Ferris, Pauline Ridge and Peter Watts for their comments and advice on this article. Thanks are also due to participants at the 'Exploring Contract Law' symposium, University of Western Ontario, many of whom furnished comments and support when the paper was first delivered. The usual caveats apply.

[1] [1923] NZLR 1106 (NZSC) 1109.

[2] *Ibid*.

[3] *Ibid*.

[4] I address *inter vivos* transactions only. I am not concerned with undue influence affecting testamentary gifts, which expresses a different principle and is governed by the probate court. For a recent discussion on the relationship between equitable undue influence and undue influence in probate law, see P Ridge, 'Equitable Undue Influence and Wills' (2004) 120 *LQR* 617.

A. Two Types of Undue Influence

At that time, and until 1985, the law relating to undue influence, though on occasion undoubtedly imperfectly expressed, was in the United Kingdom, as in other major Commonwealth jurisdictions, both conceptually tractable and (more or less) clear in terms of its criteria and purposes. One branch of the equitable jurisdiction—variously called 'actual', 'express', 'non-relational' or 'Class 1' undue influence—regulated conduct that was straightforwardly coercive, exploitative, manipulative or deceptive toward a peculiarly vulnerable party. Such conduct was unacceptable for that reason *simpliciter*. Mostly this was equity's counterpart to the common law's erstwhile, overly circumscribed duress doctrine, and it was closely related to, if not merely a manifestation of, equity's cognate jurisdiction to relieve against equitable fraud in the manner of 'unconscionable' (or 'unconscientious') dealing.[5] Thus, characteristically, the Class 1 undue influence jurisdiction regulated objectionable conduct affecting another's will, hence that was 'responsibility-relieving' from the victim's standpoint, even if it occurred outside a fiduciary relationship. The category was, therefore, distinctively non-relational, or 'relationship-independent', in the sense that it was not functionally directed at preserving the integrity of a generic class of task or relation, except to the extent that such a task or relation might have, in the particular instance, afforded an opportunity for 'actual' or 'overt' wrongdoing (for example coercion) to actualise between the parties concerned.[6] The focus of Class 1 undue influence was, accordingly, purely transactional and remedial, the public basis of intervention being the principle that 'no one shall be allowed to retain any benefit from his own fraud or wrongful act'.[7]

The other branch of the jurisdiction, in contrast—variously called 'presumed', 'relational' or 'Class 2' undue influence—was of much narrower scope and *sui generis*. It was, distinctively, a contextual application of fiduciary accountability. 'Undue influence' here possessed a rather different meaning than under the Class 1 jurisdiction. It involved a form of complaint that could only occur within a fiduciary relationship—hence its long-standing association with relations involving 'dependence and trust',

[5] *Earl of Chesterfield v Janssen* (1751) 2 Ves Sen 125, 28 ER 82 (Ch) 155–6 (Lord Hardwicke). The phrase 'undue influence' appears to have been first used by Lord Hardwicke in *Morris v Burroughs* (1737) 1 Atk 399, 26 ER 253 (Ch) 403.

[6] As occurred, for instance, in *Re Craig* (1970) [1971] 1 Ch 95. *Cf Royal Bank of Scotland plc v Etridge (No 2)* [*Etridge*] (2001) [2002] 2 AC 773 (HL) [103] (Lord Hobhouse).

[7] *Allcard v Skinner* (1887) 36 Ch D 145 (CA) 171 (Cotton LJ). *Cf CIBC Mortgages Plc v Pitt* [1994] 1 AC 200 (HL) 209 (Lord Browne-Wilkinson), explaining that the effect of the defendant's wrongdoing is to prevent the plaintiff from freely and informedly consenting to the transaction, which 'accordingly must be set aside in equity as a matter of justice'.

'trust and confidence', and the like.[8] Despite evolving and being administered as a stand-alone body of doctrine,[9] Class 2 undue influence nevertheless shared with wider fiduciary law the same generic functional purpose or substantive policy goal of controlling the narrow mischief of disloyal opportunism by those legally required to act for, on behalf of, or in the interests of another.[10] Although the jurisdiction shared with other exculpatory doctrines of the common law and equity a concern for 'fair dealing' and the quality of interpersonal transactional consent—so that the law's function in this field could in part be seen, too narrowly, as a mere corrective for one party's 'will' having been wrongfully affected by another in the formation of the impugned transaction[11]—the additional and overarching concern of the Class 2 undue influence jurisdiction was undeniably to prevent abuses of trust, as understood in conventional fiduciary terms.[12] 'Wrongfully influencing the claimant's will' was merely the mechanism by which value was diverted in a manner inconsistent with the purposes for which the parties' special relation of influence existed, or in equitable contemplation was deemed to exist.[13] The foundation for intervention in such cases was thus framed squarely by reference to 'public policy' ('trust maintenance', 'risk management', 'prophylaxis') rather than in terms of a specific 'wrongful act' having been proved against the influential party on the normal civil preponderance.[14] It followed that

[8] The relations that attract the jurisdiction have been variously described in the jurisprudence on the subject. Some courts have observed how the expressions 'relation of influence', 'relation of confidence', and 'fiduciary relation' are often used interchangeably in this context, while reminding the reader that they are not necessarily coextensive in application: see *Jenyns v Public Curator (Q)* (1953) 90 CLR 113 (HCA) 132 (Dixon CJ, McTiernan J and Kitto J). As will become clear below, special influence is legally presumed in a few well-known types of relation (eg solicitor–client, parent–child, doctor–patient, and guardian–ward). These are generally known today as 'Class 2(A)' relationships of influence. However, special influence can also exist outside of these traditional, status-based relations, and the plaintiff may prove it as a fact in the circumstances of the particular case. The label 'Class 2(B)' undue influence is commonly assigned to such cases. It is vital to bear in mind that the type of relation in question is identical as between the two sub-classes of Class 2 undue influence. The nature of the special relation or function that attracts the Class 2 jurisdiction is discussed below at text accompanying nn 72–9.

[9] Paul Finn claims that this is 'more for historical reasons than for reasons of sound principle': PD Finn, 'The Fiduciary Principle' in TG Youdan (ed), *Equity, Fiduciaries and Trusts* (Toronto, Carswell, 1989) 43.

[10] Here meaning the unauthorised diversion of the value of assets falling within the scope of the particular fiduciary function or task.

[11] See, eg, *Commercial Bank of Australia Ltd v Amadio* (1983) 151 CLR 447 (HCA) 461, where Mason J explained that relief for undue influence is given because 'the will of the innocent party is not independent and voluntary because it is overborne'.

[12] See, eg, LA Sheridan, *Fraud in Equity: a study in English and Irish law* (London, Pitman, 1957) ch 5. See also Part II below.

[13] '[A]ffecting another's "will" is merely one way in which an advantage can be taken in a relationship of ascendancy or trust': Finn, above n 9, at 45.

[14] The clearest statement of this is found in Cotton LJ's judgment in *Allcard v Skinner*, above n 7, at 171: 'In the [Class 2 undue influence] cases the Court interferes, not on the

undue influence could be presumed, and in several cases was presumed,[15] even in the absence of evidence of any actual wrongdoing or exercise of power (that is, beyond that which could be inferred provisionally from the nature of the relation and the particular transaction *inter se*).[16] Public policy dictated that relations or tasks marked by the fiduciary function should be protected from the mere possibility of abuse, including abuse in the manner of an unauthorised use of any special capacity to influence that was associated with such relations or tasks.

In practical terms the protection in the Class 2 line of cases took the form of particularised procedural rules that were designed to safeguard the integrity of fiduciary relationships, tasks or functions by managing the risk, when real,[17] against natural human tendencies to disloyalty. The most

ground that any wrongful act has in fact been committed by the donee, but on the ground of public policy, and to prevent the relations which existed between the parties and the influence arising therefrom being abused.' In my view this passage is apt to confuse, and requires explanation. Public policy justifies giving artificial effect to the evidence via a presumption of undue influence, but it is whatever is presumed thereby (undue influence) that justifies court interference, even though, at the same moment, the court must accept that undue influence may not have been proved in fact on the normal civil preponderance. In other words, public policy is the source of the presumption, but not its content. The 'presumption' is not that undue influence might have occurred (which is the reason for the presumption), but rather that it did occur (which is the reason for exculpation from the transaction) (see also n 18 below). It follows that once the presumption is properly activated, the evidence in reply must address the content of the presumption and not the (public policy) reasons for the presumption triggering in the first place. The effect of a non-displaced presumption is that, in the absence of some defence being available to a defendant, the presumption is controlling and the transaction will be set aside on the ground of 'undue influence' rather than 'public policy' per se.

[15] See, eg, *Allcard v Skinner*, above n 7, at 183 (Lindley LJ); *Johnson v Buttress* (1936) 56 CLR 113 (HCA) 133–4 (Dixon J), 143 (McTiernan J); *Hartigan v International Society for Krishna Consciousness Inc* [2002] NSWSC 810.

[16] This is not to be confused with the suggestion that a transaction can be impeached on the ground of relational undue influence even if the defendant's conduct is found to be 'unimpeachable': *cf Hammond v Osborn* [2002] EWCA Civ 885, [32] (Nourse LJ), [61] (Ward LJ). If the defendant is found not to have exercised undue influence, hence acted conflictually (that is, self-interestedly or inconsistently with his or her special duty of loyalty), it cannot follow that he or she can be presumed, counterfactually, to have done so: *cf Geffen v Goodman Estate* (1991) 81 DLR (4th) 211, [1991] SCR 353, 244 (Sopinka J). As Lord Scott pointed out in *Royal Bank of Scotland plc v Etridge (No 2)*, above n 6, at [219]: 'it makes no sense to find, on the one hand, that there was no undue influence but, on the other hand, that the presumption applies,' although his Lordship's comments are made on the basis that 'presumed' undue influence is merely a form of 'actual' undue influence proved by way of a non-contradicted permissible inference.

[17] As I have discussed elsewhere, there must be 'a conflict or a real or substantial possibility of a conflict' in order to trigger the conflict rule/presumption of undue influence: R Bigwood, 'Undue Influence in the House of Lords: Principles and Proof' (2002) 65 *MLR* 435, 447–9. The rule/presumption will not be triggered if the influential fiduciary's opportunity to pursue a personal (or a third party's) interest (a) though in itself strong enough to be an inducement, is too remote from the actual area of the fiduciary's responsibilities vis-à-vis his beneficiary, or (b) though itself sufficiently proximate to those responsibilities, is too weak an inducement realistically to be a possible determining motive on the part of the fiduciary. The subordinate party thus has no prima facie complaint, and the fiduciary will not be called on

significant of those rules involved the ritualistic implementation of a presumption of undue influence that effectively authorised a court to give artificial effect to the claimant's evidence. Being informed by the conventional fiduciary rationale, such a presumption was motivated by the strong public desire to avoid unauthorised conflicts of interest and duty by those who were required to act exclusively in the interests of another. Hence, as soon as the right combination of circumstances existed—opportunity (fiduciary influence), incentive (the impugned benefit) and epistemological uncertainty (serious detection and evidentiary problems)—equity indulged in a presumption that self-interest and temptation had operated in the particular case: that what was feared had materialised.[18] It presumed that fiduciary influence had been exercised, actively or passively, and in any event conflictually,[19] in the procurement or receipt of the impugned benefit, and it cast upon the benefiting party, if he or she wished to maintain the benefit, the burden of positively demonstrating the righteousness of the transfer or conveyance[20]—of showing that undue influence had not been exercised, but rather that the benefit received was 'the independent and well-understood act of [the plaintiff]', who had been 'in a position to exercise a free judgment based on information as full as [the defendant's own]'.[21] Put another way, 'The risk is that [the defendant] may put his own interests ahead of those of [the plaintiff]. The obligation is to prove that he has not done so.'[22] As in regular fiduciary contexts, this shifting of the

to justify or explain, where the personal opportunity taken up is, realistically, unconnected or unrelated to his or her fiduciary office or function (eg, in the negotiation of a professional fee for the fiduciary's services), or where, again realistically, the beneficiary has no demonstrable interest or expectancy in the subject matter of the transaction *inter se* (eg, trivial or moderate gifts). Indeed, such exclusions from the ambit of the conflict rule are merely acknowledgements of the purpose of the rule itself: as opportunity and temptation abate, so too does incentive, and so too does the risk of abuse (fiduciary opportunism). The influential fiduciary is accordingly free to take up the opportunity as if the relationship were an arm's-length one.

[18] As Salmond J made clear in *Brusewitz v Brown*, above n 1, at 1110: '[A relation in which undue influence occurs is] … such a relation of superiority on the one side and inferiority on the other … and therefore such an opportunity and temptation for the unconscientious abuse of the power and influence so possessed by the superior party, as to justify the legal presumption *that such an abuse actually took place* and that the transaction was procured thereby' (emphasis added). Instructive in this connection, too, is JC Shepherd, *The Law of Fiduciaries* (Toronto, Carswell, 1981) 149–50.

[19] That is to say, it was presumed that the impugned benefit was the result of the defendant acting inconsistently with the responsibility to use his or her special influential capacity exclusively in the interests of the plaintiff, who was subject to the influence and had entrusted the defendant with his or her welfare.

[20] Note that the burden was on the defendant to show affirmatively that the transaction was the plaintiff's free and properly understood act, or otherwise not the result of an abuse of his or her position. This meant that the plaintiff's independence could not be shown simply by the absence of any evidence that the defendant had exercised influence over the plaintiff. Generally see *Johnson v Buttress*, above n 15, discussed below at text below accompanying nn 84–95.

[21] *Johnson v Buttress*, *ibid*, at 134 (Dixon J).

[22] *Re P's Bill of Costs* (1982) 8 Fam LR 489, 496 (Evatt CJ and Fogarty J).

burden of proof onto the defendant to disprove undue influence—to demonstrate that there had in fact been no unauthorised conflict of interest or duty that operated to vitiate transactional consent—served to protect fiduciary relationships, tasks or functions by helping to ensure that those entrusted with the fiduciary function acted consistently with the responsibilities that such a function entails. Given the inherent difficulty of proving a negative to this effect, the application of the presumption effectively imposed strict liability, even if in theory liability for undue influence was based on 'fault' or 'blame'.[23] Again, this served the substantive goals of fiduciary law by holding those entrusted with fiduciary tasks to a higher (that is, a strict) ethic, thereby removing any possible incentive to disloyalty. It also reflected the greater importance of 'prevention' as the goal of fiduciary rules relative to the wider civil law (tort, contract, equitable wrongs such as Class 1 undue influence, estoppel and unconscionable dealing), where the paradigm is 'repair' for the doing of harm rather than the prevention of harm.[24]

The presumption of undue influence was, therefore, a pragmatic legal construction rather than 'genuinely evidential'. It was an exercise in applying, *mutatis mutandis*, the conventional 'no-conflict' rule to the parties' specific relationship and transactional encounter. It was a technique by which the law expressed and effectuated a *value judgment* about the strength of the various interests to which 'responsibility judgments' relate in the fiduciary context.[25] It was, like other policy-based presumptions, a self-conscious judicial response to societal pressures and values in connection with a particularly invidious mischief, one that transcended the simple rational relationship between the basic evidentiary facts and the presumed fact in the individual case. The evidential effect of the presumption was thus significantly greater than the natural probative weight or inferential worth of the basic facts that sufficed to generate the presumption, were those facts assessed in isolation from the substantive policy reasons for the presumption. The basic facts merely had to give rise to a 'realistic suspicion' of undue influence, as distinct from a 'legitimate inference' of it. Again, once the circumstances of actual or possible conflict between personal interest and fiduciary duty were present, equity intervened to prevent the mere possibility of self-interest being preferred to duty, in disputes where, moreover, the circumstances typically rendered proof of actual wrongdoing impossible, or at least extremely difficult.[26]

[23] *Cf* P Cane, *Responsibility in Law and Morality* (Oxford, Hart Publishing, 2002) 46 (concerning *res ipsa loquitur*), 91–2.

[24] *Cf* Cane, *ibid*, at 133.

[25] I paraphrase here a point made by Cane, *ibid*, at 91.

[26] *Cf Re Craig*, above n 6, at 104 (Ungoed-Thomas J). Such difficulty of proof exists typically because of the secret nature of the dealings between the plaintiff and the defendant, and often also because of the death of the transferor.

And although this increased the risk of misattribution of liability, and so possibly worked a hardship upon factually innocent fiduciaries in some cases, this was a risk worth tolerating given the wider interest that society has in maintaining fiduciary relations or tasks.[27] In other words, the presumption of undue influence was a significant concession to an overt public policy. It was intended to benefit claimants in a material way, by making it easier for them to succeed in their claims against influential fiduciaries and by making defences harder to maintain.[28] Even though the basic facts that triggered the presumption may not have possessed the logical value to justify a finding (via inference) of undue influence as matter of probability, public policy nevertheless validated the courts in bridging the gap and giving the logical connection between the basic facts and the presumed fact an added evidential weight that it would not ordinarily carry. Again, this was a deliberately constructed attempt, especially on the part of nineteenth-century equity judges, to manage against the pernicious risk of conflicts of interest and duty in fiduciary settings.[29]

It followed from all this that relational undue influence was 'wrongful' because, unlike what was occurring in the arm's-length Class 1 cases, and whatever else was actually involved in the individual case (coercion, misrepresentation, incapacity or the like), it entailed conflictual conduct and hence liability: it involved the unauthorised use, or presumed unauthorised use, of influence in a relationship where self-denial was both

[27] As Lamond J observed in *Bradley v Crittenden* [1932] SCR 552, [1932] 3 DLR 193, 569: 'The rule of equity which places on the donee the burden of proving both the gift and the independence of the donor's will in making it, may be a harsh one and, in individual cases, may lead to hardship. The courts, however, have found it necessary to maintain it in order to prevent those in a position to exercise undue influence from taking advantage if their position under circumstances in which proof thereof would be impossible.'

[28] In other words, although on one level the presumption could be seen merely to manage forensic difficulties, it was more than that. It incorporated a substantive judgment that those transferring suspicious benefits within the scope of a fiduciary relationship should succeed in their claim against the fiduciary if there was room for doubt as to the motives that inspired the transfer in the particular case. The risk of non-persuasion on the issue of relational undue influence shifts when the realistic possibility of such wrongdoing is shown to exist. When aided by the presumption, the plaintiff wins summarily (if the defendant cannot rebut the presumption) even though the evidence the plaintiff supplied under the burden of production would not otherwise have satisfied the burden of persuasion. The presumption assists the plaintiff in a material way, by treating him or her as if he or she had met the persuasive burden, which is purely a concession to the substantive policy that informs this area of the law.

[29] Lord Eldon put it very strongly in two judgments: in *Gibson v Jeyes* (1801) 6 Ves Jun 266 (Ch) 276, 31 ER 1044 he stated: 'It is necessary to say broadly, that those, who meddle with such transactions, take upon themselves the whole proof, that the thing is righteous. The circumstances, that pass upon such transactions, may be consistent with honest intentions: but they are so delicate in their nature, that parties must not complain of being called on to prove, they are so,' and in *Hatch v Hatch* (1804) 9 Ves Jun 292 (Ch) 297, 32 ER 615 he stated: '[I]f the Court does not watch these transactions with a jealousy almost invincible, in a great majority of cases it will lend its assistance to fraud.'

expected and required on the part of the influential actor. This also meant that it was, strictly speaking, pointless to inquire into the precise manner of the exercise of the special influence in such cases: it was unnecessary to ask what was being 'presumed' in a successful Class 2 undue influence claim. Regardless of how fiduciary influence is exercised in relational undue influence situations (pressure, flattery, advice, argument, pleading, intercession, non-disclosure, concealment of self-interest, discouraging recourse to others for advice and so on), the fact that such influence is successfully employed at all in a manner inconsistent with the defined and limited purposes of the parties' special relation affords (all else being equal) a sufficient juristic reason to suppress or reverse the impugned transaction as against the conflicted party. The general goal of fiduciary regulation is to control opportunism in limited-access arrangements, whatever form it takes. Relational undue influence, therefore, never regulated a discrete and well-defined *modus operandi* of victimisation like the law relating to other forms of objectionable conduct such as duress and misrepresentation, and so it has not lent itself to any concrete definitional approach in the manner that most other exculpatory categories have allowed. To be sure, all that is 'presumed' when the 'presumption' of undue influence operates is that the defendant has, through his or her special influential capacity over the plaintiff, and without consent or authorisation, preferred self-interest to duty, and that the plaintiff has, as a result, become the victim of misplaced trust or confidence vitiating the transactional consent.

B. The Law Changes

In 1985, this reasonably clear and stable picture of the dual strands of equity's bifurcated jurisdiction to interfere with transactions on the ground of undue influence began to dissolve, at least in the United Kingdom. Lord Scarman, on behalf of the House of Lords in *National Westminster Bank plc v Morgan*,[30] denied that the basis for relief in this area was 'a vague "public policy"'; it was, in contrast, 'specifically the victimisation of one party by the other'.[31] Accordingly, in his Lordship's view, no

> presumption of undue influence [could] arise from the evidence of the relationship of the parties without also evidence that the transaction itself was wrongful in that it constituted an advantage taken of the person subjected to the influence which, failing proof to the contrary, was explicable only on the basis that undue influence had been exercised to procure it.[32]

[30] [1985] AC 686 (HL) [*Morgan*].
[31] *Ibid*, at 705, citing Lindley LJ in *Allcard v Skinner*, above n 7, at 182–3.
[32] *Ibid*, at 704.

'Presumed undue influence' thus involved a presumption triggered by the probability (or serious likelihood) of abuse rather than the mere possibility of it. In relation to contractual transactions at least, this meant that the presumption of victimisation in the manner of relational (Class 2) undue influence could not arise unless the impugned transaction was shown to be 'manifestly disadvantageous' to the claimant, as judged by the objectively unfair terms of the impugned contract itself[33]—a stance very different from the classical fiduciary approach based on the 'realistic possibility' of abuse that had been applied by the Court of Appeal below.[34]

In 1994, obiter dicta by Lord Browne-Wilkinson in *CIBC Mortgages plc v Pitt* signaled that the traditional account of undue influence might eventually be restored.[35] His Lordship remarked that the House of Lords might need to consider 'the exact limits of *Morgan* … in the future,' and in particular the relationship, if any, between the fiduciary principle and relational undue influence.[36] However, any hope of rigorous examination of that relationship was frustrated when, in 2001, the House of Lords in *Royal Bank of Scotland plc v Etridge (No 2)*[37] failed to address (let alone restore) the fiduciary account of relational undue influence. On the contrary, and with virtually no reference to the prior decisions and academic literature on the subject, the House of Lords asserted that 'presumed undue influence' was not a distinct legal phenomenon, but rather merely an example of the way in which undue influence, typically in the manner of 'unfair persuasion', could be proven in a certain category of case involving relationships that were peculiarly vulnerable to abuse by way of non-overt acts of persuasion. Undue influence, therefore, was presented as an integrated and unitary legal phenomenon or doctrine, with the effect that it now has a universal connotation regardless of the relational context in which it occurs. In all cases the claimant bears the onus of proving undue influence in fact, so that all successful claims in this area must necessarily involve undue influence that is 'actual' in the sense that the claimant has made out his or her case against the wrongdoer on the normal civil standard.[38] However, in some cases—in particular those falling under the traditional nomenclature of 'presumed' undue influence—

[33] *Ibid*.

[34] As Slade LJ in the Court of Appeal in *National Westminster Bank plc v Morgan* [1983] 3 All ER 85 (CA) 92 recognised, 'Where a transaction has been entered into between two parties who stand in the relevant relationship to one another, it is still possible that the relationship and influence arising therefrom has been abused, even though the transaction is, on the face of it, one which, in commercial terms, provides reasonably equal benefits for both parties.' See also Dunn LJ, *ibid*, at 90.

[35] Above n 7 [*Pitt*].

[36] *Ibid*, at 209.

[37] Above n 6.

[38] See especially Lord Clyde, *ibid*, at [93].

the plaintiff succeeds by way of factual inference from naturally preponderating primary facts that were proven and insufficiently answered by counter-evidence on the other side. In other words, 'presumption' here means 'permissible inference', and the mandatory effect and the prophylactic content and function of the traditional presumption are gone. By implication, the purpose of the jurisdiction is thus purely remedial, being concerned only with responding to the genuineness of the plaintiff's transactional consent relative to the defendant's 'conscience' in the particular case, and it does not serve any higher regulatory function external to the individual exculpatory claim itself (such as prophylaxis in relation to fiduciaries generically). Also, the principle in the so-called 'relationship' cases was said in *Etridge* not to be confined to instances of 'abuse of trust and confidence'; it extends as well to arm's-length transactions—to 'cases where a vulnerable person has been exploited'.[39]

The effects of *Etridge* are obvious if not entirely explicit in the judgments. The fiduciary rationale is eviscerated, despite, as we shall see, their Lordships describing certain relationships that attract the jurisdiction in classic 'fiduciary' terminology. This explains the reference in this article's title to the 'disintegration' of undue influence in the United Kingdom: that Class 2 'relational' undue influence has been detached from its historical 'public policy' foundations and no longer has special regulative effect. But this process of disintegration has been one of 'integration' as well. At least from a forensic standpoint, Class 1 and Class 2 undue influence now possess a unitary, but still unavoidably vague, meaning. Moreover, undue influence is aligned both procedurally and (more unfortunately it seems) substantively with other exculpatory categories known to the common law and equity such as duress, misrepresentation and unconscionable dealing (at least in its Antipodean form).[40] So profound has this process of integration been in the United Kingdom that it has, in my view, become almost impossible to distinguish undue influence practically and intellectually from certain other exculpatory categories—unconscionable dealing especially[41]—such that it has lost the narrow meaning and specific function

[39] *Ibid*, at [11] (Lord Nicholls).

[40] All references herein to 'unconscionable dealing' are to the developed antipodean version of the doctrine, as expounded, eg, in *Commercial Bank of Australia Ltd v Amadio*, above n 11. For an extended discussion of that jurisdiction, see R Bigwood, *Exploitative Contracts* (Oxford, Oxford University Press, 2003) ch 6.

[41] The similarities are remarkable, especially if one accepts the 'shifting onus' approach sometimes found in judicial formulations of the unconscionable dealing doctrine, such as when judges describe an equitable 'presumption' arising from proof of 'inequality of bargaining power' and 'substantial unfairness of the bargain obtained by the stronger party'. See, eg, *Morrison v Coast Finance Ltd* (1965) 55 DLR (2d) 710 (BCCA) 713 (Davey J); *Harry v Kreutziger* (1978) 95 DLR (3d) 231 (BCCA) 237 (McIntyre J); *Commercial Bank of Australia Ltd v Amadio*, above n 11, at 474 (Deane J). It is fairly clear, though, that the legal burden of proving unconscionable dealing rests throughout on the party asserting the claim,

that it once possessed. Undue influence, therefore, risks becoming something of an equitable 'catch-all' category, much in the way that its sibling, unconscionable dealing, seems increasingly to be utilised in Australia.[42] To my mind it has become virtually impossible to understand the post-*Etridge* doctrine of undue influence in any meaningful way, so as to identify, and more importantly justify, its independent status apart from those other categories, especially unconscionable dealing.

In this article I want to describe, rather than to fully defend, the classical fiduciary account of relational undue influence. That account, I shall conclude, is defensible, but I am content for present purposes to treat it as a historical legal fact that subsisted in the United Kingdom until around 1985, and which continues to subsist elsewhere today (for example in Canada and Australia). I want then to trace, relative to that account, the progressive demise of relational undue influence as a category within the framework of fiduciary regulation, through the House of Lords decisions in *Morgan* and *Etridge* especially, and to comment on the bases for rejection of the fiduciary rationale. In *Morgan* the bases for rejection of that rationale were not developed in a convincing way, and in *Etridge*, where one might have expected concentration on the subject in the light of Lord Browne-Wilkinson's obiter dicta in *Pitt*, assertion appears to have substituted for precedent and analysis.

II. ELABORATING RELATIONAL UNDUE INFLUENCE AS A FIDUCIARY RULE[43]

From a legal history standpoint, the essential fiduciary nature of relational undue influence is undeniable. It can be traced through a respectable line of nineteenth and twentieth-century authorities,[44] in addition to Dixon J's

and that 'presumption' here means 'permissible inference', with a shift in the evidential onus only: see, eg, *Diprose v Louth* (1992) 175 CLR 621 (HCA) 632; *Micarone v Perpetual Trustees Australia Ltd* (1999) 75 SASR 1 (SC) 127; *Smyth v Szep* [1992] WWR 673 (BCCA) 681–2; *Gindis v Brisbourne* (2000) 183 DLR (4th) 431 (BCCA) 442–3.

[42] *ANZ Banking Group Ltd v Karam* (2005) 64 NSWLR 149 (CA) where the court recommended absorption of lawful-act duress into unconscionable dealing. It might be argued that unconscionable dealing is better suited to regulating unfair persuasion in arm's-length transactions than relational undue influence.

[43] Material in some subsections of this section has been reproduced, though mostly in an edited form, from Bigwood, above n 40, ch 8.

[44] See, eg, *Gibson v Jeyes*, above n 29, at 1049–50; *Huguenin v Baseley* (1807) 14 Ves Jun 273, 33 ER 526 (Ch) 531–532; *Billage v Southee* (1852) 9 Hare 534 (Ch) 540, 68 ER 623; *Wright v Vanderplank* (1856) 8 De GM & G 133 (Ch) 137, 44 ER 340; *Tate v Williamson* (1866) 2 Ch App 55, 60; *Allcard v Skinner*, above n 7, at 171; *Bradley v Crittenden*, above n 27, at 559; *Tufton v Sperni* [1952] 2 TLR 516 (CA) 522 (Sir Raymond Evershed MR), 530 (Jenkins LJ); *Lloyds Bank Ltd v Bundy* [1975] QB 326, especially the judgment of Sir Eric Sachs; *O'Sullivan v Management Agency and Music Ltd* [1985] QB 428,

exemplary account of the subject in *Johnson v Buttress*,[45] which will be discussed further.[46] There was, and indeed still is, significant academic opinion in support of the fiduciary personality of relational undue influence in English-based legal systems, including the United Kingdom, Australia, Canada and the United States.[47] Understanding undue influence as a category within the framework of fiduciary regulation, however, can be made difficult by reliance on apparent, and often non-conceded or unappreciated, misunderstandings and misdescriptions of the conventional boundaries of fiduciary accountability by modern courts and commentators.[48] Indeed, as will be seen below, virtually all denials of the essential fiduciary nature of relational undue influence law involve a basic failure to explore, appreciate or respect the conventional boundaries of fiduciary obligation and regulation. Below I set out my own understanding of those boundaries, with a view to showing how relational undue influence rests naturally within them.[49]

448–9. For an explicit acknowledgement, after *Morgan*, of the substantial identity between undue influence and breach of fiduciary duty, see *Goldsworthy v Brickell* [1987] Ch 378, 400–401 (Nourse LJ).

[45] Above n 15.

[46] See text below accompanying nn 84–95.

[47] Examples include S Williston, *Williston on Contracts* (New York, Baker, Voorhis, 1937) § 1625ff; Sheridan, above n 12, at ch 5; LS Sealy, 'Fiduciary Relationships' [1962] *CLJ* 69, 78; PD Finn, *Fiduciary Obligations* (Sydney, Law Book Company Ltd, 1977) 41ff; Shepherd, above n 18, at ch 14; J Glover, *Commercial Equity: Fiduciary Relationships* (Sydney, Butterworths, 1995) 8ff, 33; AJ Duggan, 'Undue Influence' in P Parkinson (ed), *The Principles of Equity*, 2nd edn (Sydney, Law Book Company Ltd, 2003) 428–31; R Flannigan, 'The Fiduciary Obligation' (1989) 9 *OJLS* 285; PD Maddaugh and JD McCamus, *The Law of Restitution*, 2nd edn (Toronto, Canada Law Book, 2004) 812; G Ferris, 'Why is the Law of Undue Influence so Hard to Understand and Apply?' in E Cooke (ed), *Modern Studies in Property Law—Volume 4* (Oxford, Hart Publishing, 2007) ch 3. Ashburner viewed the presumption of undue influence, as a ground for rescinding contracts, as 'so closely connected with the doctrines [relating to conflict of interest and duty]' that he dealt with them in a separate chapter dedicated to that subject: see D Browne, *Ashburner's Principles of Equity*, 2nd edn (London, Butterworths, 1933) 299, ch 21.

[48] As to which see generally R Flannigan, 'The Boundaries of Fiduciary Accountability' (2004) 83 *Canadian Bar Review* 35, [2004] *New Zealand Law Review* 215. The references herein are to the *New Zealand Law Review* reprint of the article.

[49] As the following text and footnotes reveal, my understanding of the conventional boundaries of fiduciary accountability owes a large debt to Robert Flannigan's lucid and sustained writings in the field. These include Flannigan, above n 47; R Flannigan, 'Commercial Fiduciary Obligation' (1998) 36 *Alberta Law Review* 905; R Flannigan, 'Fiduciary Regulation of Sexual Exploitation' (2000) 79 *Canadian Bar Review* 301; Flannigan, above n 48; R Flannigan, 'Fiduciary Duties of Shareholders and Directors' [2004] *Journal of Business Law* 277; R Flannigan, 'A Romantic Conception of Fiduciary Obligation' (2005) 84 *Canadian Bar Review* 391; R Flannigan, 'The Adulteration of Fiduciary Doctrine in Corporate Law' (2006) 122 *LQR* 449; R Flannigan, 'The Strict Character of Fiduciary Liability' [2006] *New Zealand Law Review* 209; R Flannigan, 'The Economics of Fiduciary Accountability' (2007) 32 *Delaware Journal of Corporate Law* 393; R Flannigan, 'The [Fiduciary] Duty of Fidelity' (2008) 124 *LQR* 274.

A. Fiduciary Regulation in a Nutshell

Fiduciary regulation exists exclusively to ensure selfless loyalty by those who are required to act for, on behalf of, or in the interests of, another. It serves no other juridical purpose.[50] Although a fiduciary obligation is imposed[51] on certain arrangements by private law, the basic justification for recognising and regulating relationships or tasks of a fiduciary nature is essentially a 'public' or 'social' one.[52] It is 'informed by considerations of public policy aimed at preserving the integrity and utility of [fiduciary] relationships [or tasks] given the expectation that the community is considered to have of behaviour in them, and given the purposes they serve in society'.[53] It is also justified by the significant opportunities that exist for fiduciaries as a class to act inconsistently with the equitable purposes of their office or function (the risk of the mischief sought to be suppressed by the imposition of fiduciary obligation is serious and high) in combination with the monitoring and evidentiary problems that characteristically attend

[50] Nominate duties on the part of fiduciaries that go beyond exacting loyalty (eg, duties of care), or defalcations by fiduciaries that do not involve disloyalty per se (eg, discretionary unequal distributions to beneficiaries) are not fiduciary duties or defalcations, and hence do not attract fiduciary accountability or liability. According to Flannigan, fiduciary accountability 'is concerned exclusively with controlling opportunism on the part of those with limited access. It is not concerned with defining substantive performance standards for nominate functions. Nor is it concerned with lack of care, bad judgment, unjust enrichment, power differentials, general market exploitation or the substantive merits of decisions': R Flannigan, 'Fiduciary Duties of Shareholders and Directors,' above n 49, at 279. See also S Worthington, 'Fiduciaries: When Is Self-Denial Obligatory?' [1999] *CLJ* 500, 501–503. We must therefore distinguish between fiduciary obligations and the full range of idiosyncratic or nominate responsibilities that may be owed by a fiduciary to his or her obligee. For judicial recognition of this point, see *Pilmer v Duke Group Ltd (in liq)* (2001) 207 CLR 165 (HCA) 196, 217 (Kirby J), citing *Norberg v Wynrib* (1992) 92 DLR (4th) 449, [1992] 2 SCR 226, 272 (McLachlin J); *Bristol and West Building Society v Mothew* (1996) [1998] Ch 1, 16 (Millett LJ); *Henderson v Merrett Syndicates Ltd* (1994) [1995] 2 AC 145 (HL) 206 (Lord Browne-Wilkinson).

[51] At root, fiduciary status and thus accountability is an 'imposed' form of obligation rather than an 'assumed' one. Although the physical arrangement that led to one party having defined and limited access to another party's assets may have been the result of a voluntary decision on the part of the parties involved (eg via contract or unilateral declaration of trust), it is the fact of the resultant limited access, and the risk of opportunism that flows from that access, rather than the consent to the physical arrangement itself that leads to the imposition of fiduciary obligation within the scope of that arrangement. *Cf* Finn, above n 9, at 46–7, 54; R Flannigan, 'Fiduciary Regulation of Sexual Exploitation,' above n 49, at 302; R Flannigan, above n 48, at 219: 'There is no condition of liability that actors subjectively agree to this kind of legal responsibility. It is only necessary that their arrangements, however created, involve limited access.'

[52] *Cf Hodgkinson v Simms* (1994) 117 DLR (4th) 161, 186, [1994] 3 SCR 377 (La Forest J).

[53] Finn, above n 9, at 42. See also R Flannigan, 'The Fiduciary Obligation,' above n 47, at 321–2; *Chirnside v Fay* [2004] 3 NZLR 637 (CA) 51.

the decisions and actions of those holding such offices or functions (the chance of detection of actual fiduciary breaches in individual cases is low).[54]

In practical terms, fiduciary loyalty is exacted through proscription of its opposite.[55] Loyalty is ensured through the operation of overlapping default rules—in particular the 'no-profit' and 'no-conflict' rules—that prohibit and seek to deter fiduciary disloyalty. The fiduciary is simply instructed to forgo self-interest while acting in the interests of another, at least without securing, *ex ante* or *ex post*, the genuine consent of the one to whom the duty of loyalty is owed.[56] In the absence of clear proof of such consent (or other lawful authorisation)—and it is for the fiduciary positively to show that such consent (or authorisation) was fully and freely obtained rather than for the obligee to show that it was not—liability is strict. There are no excuses besides consent or authorisation for disloyal actions or tendencies within the scope of the fiduciary function. Indeed, the special accountability regime that responds to breaches of fiduciary obligation—strict liability, followed by the potential availability of significant remedies such as equitable compensation, constructive trust, and account of profits—serves as a legally self-conscious attempt to suppress or deter fiduciary disloyalty by removing any conceivable incentive for the fiduciary to act inconsistently with his or her special duty of abstinence.[57] Moreover, so strong is the prophylactic stance here that a mere disloyal tendency in dealings between a fiduciary and his or her obligee, if it falls within the scope of the fiduciary's special function, is regarded as sufficient to justify a presumption of abuse and the resultant shifting of the burden onto the fiduciary to show the propriety of the impugned dealing as a whole. In other words, the potential of fiduciary liability is triggered by contingencies—a level of risk—that would indicate to a court the (realistic) possibility of abuse of fiduciary office or function rather than the probability (or substantial likelihood) of it.[58] Failing clear proof of consent or authorisation, therefore, liability might still be imposed upon a fiduciary even in the absence of

[54] As Flannigan, 'The Strict Character of Fiduciary Liability,' above n 49, at 212, comments, 'These considerations ground our insistence on a strict liability. Our determination is to reduce the prospect that fiduciaries will conclude that there are opportunities to avoid or inhibit detection, or to construct plausible explanations after the fact.' As Flannigan explains, fiduciaries are often well placed to erect plausible facades of righteousness in relation to transactions entered into with their obligees. The law's approach to the regulation of fiduciary tasks or functions must thus be sensitive and responsive to this fact or risk.

[55] *Breen v Williams* (1996) 186 CLR 71 (HCA) 113 (Gaudron and McHugh JJ); see also Gummow J, *ibid*, at 137–8; Finn, above n 9, at 2, 28.

[56] For a discussion of the defence of consent in the fiduciary context, see J Payne, 'Consent' in P Birks and A Pretto (eds), *Breach of Trust* (Oxford, Hart Publishing, 2002) ch 10.

[57] *Cf Warman International Ltd v Dwyer* (1995) 182 CLR 544 (HCA) 557–8.

[58] The difference between 'possibility' regulation and 'probability' regulation is nicely put by P Finn, *Integrity in Government (Second Report)—Abuse of Official Trust: Conflict of*

direct proof of the actual influence of the conflict in the particular case, but this is simply a concession to the strong public desire to control the very mischief and epistemological difficulties that justify imposition of fiduciary responsibility in the first place.

B. What Identifies the Imposition of a Fiduciary Obligation? The 'Limited-Access' Abstraction

Courts and commentators have long struggled to articulate a clear and universal test for identifying a fiduciary obligation in particular instances, especially when the obligation recognised is 'fact-based' rather than 'status-based'. Following Robert Flannigan's excellent work in this field, though, it can be seen that the common physical characteristic of those arrangements in which one party will be a fiduciary relative to another is 'limited access to assets': one party has, wholly or partly, acquired access to the assets of another for a 'defined or limited' purpose, that is, for the purpose of furthering the objectives of that other party to the exclusion of the first party's (unauthorised, inconsistent) personal interests. In other words, access to the other's assets is not 'open' as it is, for example, in simple exchange relations or encounters but rather 'functional and constrained' by the relevant interests of that other party (or possibly the parties' joint interests). The law recognises the serious mischief associated with limited-access arrangements—that the value of the assets can readily be diverted or exploited for self-serving ends—and it seeks to control it by imposing, to the extent of the access, 'fiduciary' responsibility.[59]

It is to be noted that the concepts of 'access' and 'assets' in this connection are liberal. 'Access to assets' refers to access that is either direct (for example, where a person is a fiduciary because he or she is holding title over another's property in a managerial capacity) or indirect (for example, where a person has become a disinterested adviser for another and so has special access to the other's assets via his or her 'will' or legal capacity for decision-making).[60] The concept of 'indirect access to assets'

Interest and Related Matters (Canberra, Australian National University, 1993) 12–13, thus: '[Possibility regulation is] … activated characteristically by the mere existence of a particular set of stipulated conditions, irrespective of whether there are actual grounds to suspect in a given instance that those conditions will, or are likely to, result in an abuse of office. [In contrast, probability regulation] … ordinarily attracts the additional requirement of actual grounds for suspicion in any given case. The practical difference between the two is that the former produces rules which in effect presume the very phenomenon of which the latter requires positive proof in some measure, ie the likelihood of an abuse in a given instance.'

[59] See generally the writings of Flannigan, above n 49.

[60] Flannigan, 'The Fiduciary Obligation,' above n 47, at 308–309.

thus epitomises the relational undue influence situation, where the defendant's transactions with the plaintiff are constrained by fiduciary responsibilities (disabilities) precisely because of the degree to which he or she knowingly controls, or is capable of controlling, the will, and hence the decision-making, of the plaintiff. The meaning of 'assets' is also expansive, because it comprehends the acquisition of intangibles such as contractual rights, which the defendant may acquire independently of the tangible assets or benefits that may flow from the enforcement of contractual rights.[61]

C. Toward an Understanding of Relational Undue Influence Inside 'Fiduciary Law': Defining the 'Fiduciary Trust' Types

Most seem to accept that the concept of 'trust' is somehow central to a fiduciary obligation. There is lesser agreement on other descriptors that have been employed from time to time in search of the defining characteristic or chief determinant of fiduciary status: 'reliance', 'vulnerability',[62] 'reasonable (or legitimate) expectations'[63] and the like. The fact is, though, that all these descriptors are unstable to the extent that each is susceptible of overly expansive interpretations.[64] For example, to say for present purposes that 'P trusts D' is unilluminating until we first understand precisely the way in which P trusts D, and for what purposes. P must trust D in some relevant way.[65] Only trust that leads to or results from D having 'limited access' to P's assets—access for P's objectives rather than for D's own—attracts a 'fiduciary' obligation, hence fiduciary regulation. On this view, moreover, fiduciary trust can only ever be identified by its consequences—fiduciary access—and the serious risk of opportunism that such access affords. Although the existence of subjective trust or confidence reposed in an alleged fiduciary can indicate that fiduciary access exists in the context of the relationship or dealing under examination, it is clear that subjective trust (or expectation) is neither a necessary nor a sufficient condition of fiduciary status. There is a significant prescriptive

[61] *Ibid*, at 308, fn 118.

[62] See, eg, *Hospital Products Ltd v United States Surgical Corporation* (1984) 156 CLR 41 (HCA) 142 (Dawson J). For an account of the Supreme Court of Canada's preoccupation with 'vulnerability' as the key defining characteristic of the fiduciary relationship, see MH Ogilvie, 'Fiduciary Obligations in Canada: From Concept to Principle' [1995] *Journal of Business Law* 638.

[63] Finn, above n 9, at 46–7; accepted by Kirby P in *Breen v Williams* (1994) 35 NSWLR 522 (SC) 544.

[64] *Cf* the comments of Kirby J in *Pilmer v Duke Group Ltd (in liq)*, above n 50, at 288.

[65] *Cf* Flannigan, 'The Fiduciary Obligation,' above n 47, at 302: 'Everyone is a fiduciary who is trusted by another in the relevant way.' *Cf* also *Breen v Williams*, above n 55, at 93 (Dawson and Toohey JJ), where this point is well recognised in relation to the doctor–patient relationship.

dimension to the concept of trust in this field of law, and, regardless of the actual belief of the parties involved, limited access either exists in connection with the particular task or relation, or class of tasks or relations, or it does not.[66] Again, this is just to underscore the fact that the 'limited-access' abstraction, rather than 'trust' per se, is the singular defining characteristic or prime determinant of fiduciary status.

Accepting that fiduciaries are 'trusted' because they have qualified (functional and constrained rather than open) access to their obligee's assets, it next remains to distinguish between the different types of 'fiduciary trust' (or 'limited access') that might exist in connection with a 'trusted' person for the purpose of attracting fiduciary accountability. Such differentiation, moreover, allows us to appreciate the sheer diversity in the range of relationships or tasks that are vulnerable to the mischief associated with limited access, and hence that are subject to the imposition of a fiduciary obligation. Its effect is to demonstrate how 'fiduciary regulation' has a much wider application in our social relations or functions than some judges and writers seem to appreciate or care to acknowledge,[67] as well as to show that the 'influential' party in a so-called 'relation of influence,' for the purposes of relational undue influence law, is invested with fiduciary access and hence is subject to the usual disabilities that are engrafted upon the transactional liberty of a self-serving fiduciary.

As Flannigan has explained, fiduciary trust may present as either of two basic types: 'vigilant' trust or 'deferential' trust.[68] Vigilant trust denotes the type of trust that is normally reposed in intermediaries[69]—in those who have an intermediary function in the employment of the trusting party's assets, which function gives the intermediary 'limited access' to those assets. It is, accordingly, of little concern to us in connection with relational undue influence liability. Trustees, solicitors, agents, directors, partners and (certain) employees are examples of those in whom vigilant trust is typically reposed by (respectively) beneficiaries, clients, principals, companies, fellow partners and employers. The fiduciary obligation is imposed

[66] Again *cf* the comments of Kirby J in *Pilmer v Duke Group Ltd (in liq)*, above n 50, at 288.

[67] Flannigan, above n 48, at 216, fn 1: 'Most of us are accountable as fiduciaries in one or more respects most of the time.' And see R Flannigan, 'Fiduciary Mechanics' (2008) 14 *Canadian Labour and Employment Law Journal* 25.

[68] In the High Court of Australia, in *Breen v Williams*, above n 55, at 82, Brennan CJ recognised a distinction between fiduciary trust types that roughly parallels Flannigan's. According to the Chief Justice, fiduciary duties arise from two distinct but frequently overlapping sources: (a) agency, and (b) relationships of ascendancy or influence by one party over another or dependence or trust on the part of that other.

[69] This label is chosen by Flannigan because those who repose 'vigilant trust' in others tend to remain wary or 'vigilant' that their trust may at any time be misplaced. Employers in respect of their servants or agents, for example, are entitled to expect that the latter will serve the former faithfully, although unreserved or unsceptical faith is rarely conceded as it is in the 'deferential trust' cases. See Flannigan, 'The Fiduciary Obligation,' above n 47, at 287.

here on the basis of the 'intermediary [or 'agency'] cost mischief'[70]—the 'cost' to the trusting party of the mischief that arises when beneficial ownership is separated from management and control.[71] The fiduciary obligation is thus imposed in an attempt to avoid those costs: that is, in order to deter opportunistic conduct.

The basis for recognising (imposing) a fiduciary obligation in cases that involve deferential trust is rather different from the intermediary cost justification that explains such an obligation in the vigilant trust cases. According to Flannigan, deferential trust is 'deferential' in the sense that 'the trusting person will defer to the judgment of the trusted person'.[72] Sometimes the deference involved is total, while at other times it is merely partial or situational.[73] In some cases such as *Johnson v Buttress*,[74] to be discussed below, it is attended by elements of necessity, dependence or submission. But in other cases no such obvious vulnerability exists, yet still the ascendant party knows that his or her impartial judgment and guidance is being relied on in the circumstances.[75] In every case it is necessary to examine the relationship between the parties involved, and in particular the relative positions and precise roles within that relationship. To this end, one party's role or position must be such as to 'implicate [him] in the other's affairs or so align him with the protection or advancement of that other's interests that foundation exists for the "fiduciary expectation."'[76] Such a role may exist, or be taken to have arisen, in a variety of ways and in a variety of settings,[77] but typically it is found 'either because [the

[70] 'Intermediary' or 'agency' costs are defined in different ways by different commentators. See Flannigan, 'The Economics of Fiduciary Accountability,' above n 49, at 397, fn 12. For Flannigan, 'agency costs are the costs of opportunism and the costs of controlling opportunism': *ibid*.

[71] See Flannigan, *ibid*, at 394–401. See also Duggan, above n 47, at 429–31.

[72] Flannigan, 'The Fiduciary Obligation,' above n 47, at 286. *Cf* also the concept of 'reverence' in Continental law: J du Plessis and R Zimmermann, 'The Relevance of Reverence: Undue Influence Civilian Style' (2003) 10 *Maastricht Journal of European and Comparative Law* 345.

[73] *Ibid*.

[74] Above n 15.

[75] Flannigan cites as examples *Huguenin v Baseley*, above n 44, at 293–4 (Lord Eldon LC) and *Lloyds Bank Ltd v Bundy*, above n 44, at 341 (Sir Eric Sachs): see Flannigan, 'The Fiduciary Obligation,' above n 47, at 287, fn 18. Commentators have fairly doubted whether *Lloyds Bank Ltd v Bundy* can be supported on a fiduciary (*qua* relational undue influence) basis. See, eg, P Finn, 'Contract and the Fiduciary Principle' (1989) 12 *University of New South Wales Law Journal* 76, 96, arguing that the case is supportable on an unconscionable dealing basis only. Eisenberg also regards this as an unconscionability (*qua* 'transactional incapacity') case: 'the real vice of the transaction was that the bank had led a relatively unsophisticated person into a transaction with a severe though not readily apparent potential for unfairness, without pointing out the need for expert advice': see MA Eisenberg, 'The Bargain Principle and Its Limits' (1982) 95 *Harvard Law Review* 741, 767.

[76] Finn, above n 9, at 47.

[77] For example, it may exist or arise by nature, independent of human will, such as in the parental or quasi-parental function, or it may have been brought about by the voluntary

defendant] is or has become an adviser of [the plaintiff] or because he has been entrusted with the management of [the plaintiff's] affairs or everyday needs or for some other reason,' and so 'is in a position to influence [the plaintiff] into effecting the transaction of which complaint is later made'.[78] The effect of the parties' relation or circumstances is always 'deference' in the sense that what follows is 'relaxation of [the plaintiff]'s self-interested vigilance or independent judgment in favour of [the defendant's] protection or judgment because the circumstances of the relationship justify the belief that [the defendant] is acting or will act in … [the plaintiff]'s interests'.[79]

Deferential trust, then, is trust that gives the trusted party special ability to influence the trusting party's decisions: it gives the former indirect access to the latter's assets, which is access through the latter's 'will' or capacity to make decisions. The former effectively has control over the latter's decision-making in connection with matters that fall within the scope of the parties' deferentially trusting relationship or encounter. Deferential trust differs from vigilant trust only in the manner in which it arises and subsequently operates as between the parties involved. Vigilant trust generally arises because the plaintiff has made a decision that he or she expects the defendant as an intermediary to implement (an 'implementational' trust, says Flannigan[80]), whereas deferential trust arises because the defendant is, or can be taken to be, serving as a disinterested adviser or confidant to the plaintiff, or because the defendant knowingly occupies a position of authority, respect or expertise relative to the plaintiff.[81] Vigilant trust thus operates in association with the implementation of decisions by the defendant as the trusted party, but it does not necessarily comprehend the defendant having an opportunity to participate in the original decision-making (by or on behalf of the plaintiff) that gave rise to the implementational trust itself. Deferential trust, in contrast, envisages input by the defendant at the decision-making stage, that is, at the moment when the defendant enters into a beneficial transaction with the plaintiff.[82] Although at that moment it may appear to an objective observer that the plaintiff is making his or her own decisions, in reality it is, or might be, the defendant's will that is operative.

Note that although vigilant trust and deferential trust differ in these important respects, of which more below, they are otherwise identical in

action of the parties themselves, such as in the professional relations of solicitor and client, or it may have originated by circumstances and conduct generally, such as in *Johnson v Buttress* (discussed below at text accompanying nn 84–95).

[78] *Goldsworthy v Brickell*, above n 44, at 401 (Nourse LJ).

[79] Finn, above n 9, at 48.

[80] Flannigan, 'The Fiduciary Obligation,' above n 47, at 293.

[81] *Ibid*, at 292–3.

[82] *Ibid*, at 293. Flannigan calls this 'decisional' trust, in contrast to 'implementational' trust.

their essential 'trust' quality for the purpose of attracting 'fiduciary' regulation in respect of certain dealings between the plaintiff and the defendant, or between the defendant and a third party. This is because recognition of a fiduciary obligation in both cases is an attempt to avoid the self-same mischief of fiduciary opportunism in limited-access arrangements. The risk in both situations is that the defendant's commitment to the defined and limited purpose of the relationship with the plaintiff will be compromised by the prospect of personal gain to the defendant (or to a third person at the defendant's direction). Although the source of the opportunity for abuse might differ as between the two forms of fiduciary trust—the separation of asset ownership from management and control in the case of vigilant trust, and the special capacity to influence the trusting party's decision-making in the case of deferential trust—the opportunities for such abuse are nevertheless identical:

> Both types of trust in fact result in the trusted party acquiring 'access' to the employment of assets. In the case of deferential trust, however, the access is indirect because it occurs through 'influence' exerted by the trusted party. But in either case, and to the same extent, the 'access' to assets may be turned to mischievous ends.[83]

D. A Consummate Judicial Encapsulation of Relational Undue Influence as 'Fiduciary' Regulation: Dixon J in *Johnson v Buttress*

Johnson v Buttress was decided by the High Court of Australia in 1936.[84] The respondent's father, Buttress, had made to the appellant, Mrs Johnson, an *inter vivos* gift of a block of land on which stood a cottage. After Buttress's death, the respondent challenged the gift on the ground of undue influence. At the time of the gift Buttress was of advanced years and recently widowed. He was illiterate and peculiarly dependent upon others. Mrs Johnson, who was a distant relative of Buttress's wife and a long-time friend of the couple, had remained close to Buttress after the wife's death and generally took care of him. The majority of the court found that Buttress, by reason of his age, illiteracy, temperament, inferior mental faculties and particular mode of life, had come to depend on Mrs Johnston completely for her guidance, support, supervision and kindness.[85] Dixon J, in particular, held that the circumstances showed that Mrs Johnson had, in

[83] *Ibid*, at 309.
[84] Above n 15.
[85] Latham CJ, Dixon J and McTiernan J delivered separate judgments to the same effect. Evatt J agreed with Dixon J's judgment. Starke J, in contrast, had difficulty in accepting that the facts disclosed a peculiar relation of trust and confidence between the donor and Mrs Johnson so as to bring him within the 'protected class' in respect of which there is a presumption of undue influence. Nonetheless, Starke J thought that the facts afforded

the time leading up to the impugned gift, in fact stood in an 'antecedent relation of influence' to Buttress. Although there was in his view no direct proof that the transaction was procured by the improper exercise of an actual ascendancy or domination gained by Mrs Johnson over Buttress, the fact that a substantial gift had been made by Buttress to Mrs Johnson within the scope of the parties' special antecedent relation of influence imposed on the latter the 'burden of justifying the transfer by showing that it was the result of the free exercise of [Buttress's] independent will'—a burden that, in Dixon J's view, Mrs Johnson had 'quite failed' to discharge.[86]

For Dixon J, the foundation of equity's jurisdiction to rescind a transfer of property for undue influence lay in 'the prevention of an unconscientious use of any special capacity or opportunity that may exist or arise of affecting the alienor's will or freedom of judgment in reference to such a matter'.[87] As was orthodoxy at that time, and indeed is still now after *Etridge*, Dixon J divided undue influence into two distinct categories according to the source of the power to exercise control over (that is, to 'influence') the alienor's will or freedom of judgment. The first category encompassed those cases where the source of the power lay in 'no antecedent relation but in a particular situation, or in the deliberate contrivance of the dominant party,'[88] although such a power may coincidentally arise in the context of such an antecedent relation. Cases falling within this category would today be recognised as Class 1, 'actual' undue influence claims, in which the party who could establish such a power or opportunity and who wished to upset the transaction would bear the onus of adducing further evidence to show that the transaction was the outcome of the influence having actually been exerted over the mind or will of the subordinate party.

The second category of undue influence identified by Dixon J encompassed those cases where the source of the power to exercise control over the alienor's will or freedom of judgment lay distinctively in the situation where 'the parties … antecedently stand in a relation that gives to one an authority or influence over the other from the abuse of which it is proper that he should be protected'.[89] Where the parties stood in such a relation, the influential party who received 'property of substantial value', for example by way of gift, from the other party could not retain the beneficial title to the subject matter of the gift 'unless he satisfie[d] the court that he

sufficient evidence to allow a court justly to infer that the transfer was involuntary as having resulted from 'unfair and undue pressure' on the part of Mrs Johnson.

[86] *Johnson v Buttress*, above n 15, at 138. *Cf* McTiernan J, *ibid*, at 143.

[87] *Ibid*, at 134.

[88] *Ibid*.

[89] *Ibid*.

took no advantage of the donor, but that the gift was the independent and well-understood act of a man in a position to exercise a free judgment based on information as full as that of the donee'.[90] Such a burden fell upon the one who received substantial benefits within the context of certain well-known relationships that 'by their very nature import influence'. (These would be known today as 'Class 2(A)' relationships of influence.) By way of example Dixon J mentioned the familiar relations of solicitor–client, parent–child, physician–patient, guardian–ward, and fiancé–fiancée, but he was also quick to indicate that the burden of justifying the transaction was not confined to fixed categories. Rather, it rested upon a principle, and the fiduciary principle no less:

> It applies whenever one party occupies or assumes towards another a position naturally involving an ascendancy or influence over that other, or a dependence or trust on his part. One occupying such a position falls under a duty in which fiduciary characteristics may be seen. It is his duty to use his position of influence in the interest of no one but the man who is governed by his judgment, gives him his dependence and entrusts him with his welfare.[91]

When such a party took a substantial gift of property from the dependent or trusting party, Dixon J explained, or entered into a transaction that 'wears the appearance of a business dealing', a presumption against the validity of the disposition arises in the latter party's favour, so that it becomes 'incumbent upon [the transferee] to show that it cannot be ascribed to the inequality between them which must arise from his special position'.[92] Such a presumption, in Dixon J's view, rested upon a 'firm foundation' of policy, and upon other considerations that combined to justify strict fiduciary regulation, namely:[93]

— that the transferee may be taken to possess a peculiar knowledge not only of the disposition itself but also of the circumstances that should affect its validity;
— that the transferee chose to accept a benefit that may well proceed from an abuse of the authority conceded to him or her, or the authority reposed in him or her; and
— that the relations between the transferor and the transferee are so close as to render it difficult to disentangle the inducements that led to the transaction.

Justice Dixon also emphasised that this rule must not be narrowed: 'the risk must not be run of fettering the exercise of the jurisdiction by an

[90] *Ibid*.
[91] *Ibid*, at 134–5.
[92] *Ibid*, at 135.
[93] *Ibid*.

enumeration of persons against whom it should be exercised; the relief stands upon the general principle applying to all the variety of relations in which dominion may be exercised by one person over another.'[94]

To my mind, Dixon J's judgment in *Johnson v Buttress* is the most pellucid judicial elaboration of the essential 'fiduciary' nature of relational undue influence that exists in the law reports of the major British Commonwealth legal systems.[95] It signifies that, in purely functional terms, the equitable jurisdiction to rescind transactions on the basis of Class 2 undue influence is indistinguishable from all other categories (such as breach of confidence and wider fiduciary obligation) that are united by the singular and generic goal of controlling opportunism in limited-access arrangements. Separate doctrinal administration by the courts should not be taken to imply that different conceptual or substantive concerns are necessarily involved in each situation. The interests protected, and the broad mischief sought to be avoided, are identical as between Class 2 undue influence and wider fiduciary law. Both fall squarely within the ambit of the same policy goal of maintaining the integrity of 'trusting' relations or institutions; both are 'fiduciary' for the same reason, that is, because one party has acquired 'limited access' to another's assets, which raises a singular opportunism concern, and which in turn justifies the imposition of 'fiduciary' accountability to the extent of the limited access. The mechanistic difference with relational undue influence is the means by

[94] *Ibid*, at 136, clearly echoing the words of Chelmsford LC in *Tate v Williamson*, above n 44, at 61.

[95] Sir Eric Sachs' judgment in *Lloyds Bank Ltd v Bundy*, above n 44, and Dunn and Slade LJJ's judgments in *Morgan*, above n 30, in the English Court of Appeal are also instructive. Although the law of undue influence appears to differ from state to state in the United States, attention should be paid to *Martinelli v Bridgeport Roman Catholic Diocesan Corp* 196 F 3d 409 (2d Cir 1999), in which the court, applying Connecticut law, observed that the conventional view is that the normal rule (that a person alleging a wrong must prove it) 'is somewhat relaxed in cases where a fiduciary relation exists between the parties to a transaction or contract, and where one has a dominant or controlling force or influence over the other. In such cases, if the superior party obtains a possible benefit, equity raises a presumption against the validity of the transaction or contract, and casts upon such party the burden of proving fairness, honesty, and integrity in the transaction or contract … Therefore, it is only when the confidential relation is shown together with suspicious circumstances, or where there is a transaction, contract, or transfer between persons in a confidential or fiduciary relationship, and where the dominant party if the beneficiary of the transaction, contract, or transfer, that the burden shifts to the fiduciary to prove fair dealing. A fiduciary seeking to profit by a transaction with the one who confided in him had the burden of showing that he has not taken advantage of his influence or knowledge and that the arrangement is fair and conscientious': *ibid*, at 421, fn 6. The court also indicated that 'Shifting the burden of proof protects fiduciary relationships by helping to ensure that the fiduciary acts consistently with the responsibilities such relationships entail': *ibid*, at 421. It did note that 'To be sure, where the fiduciary has not received some kind of benefit that would engender suspicion and there is no other evidence of wrongdoing, the burden of proof remains on the plaintiff', *ibid*. As to the proof required, the court held that 'as in other instances in which the burden shifts to the fiduciary to show fair dealing, such proof must be by "clear and convincing evidence"', *ibid*, at 423.

which the fiduciary's office or function is, or can be, misused. In a successful relational undue influence claim, it is the ascendant party's 'influence' over the deferentially trusting party, rather than a direct diversion of the value of assets, that is the mechanism of the opportunism that lies at the foundation of the latter party's claim for exculpation from the impugned transaction.

E. Respecting Administrative Differences: On Aligning 'Obligation' with 'Factual Structure'

The last point in the preceding section is a vital one, for although fiduciary obligation and relational undue influence serve in common the elemental function of controlling opportunism in limited-access arrangements, it is necessary nonetheless to observe important factual and practical distinctions that exist in connection with the administration of the respective bodies of doctrine. These distinctions, however, merely symbolise contextual regulation of an otherwise conceptually unified subject matter. They are directed not at the nature of the obligation, regulation and liabilities involved, as those are identical across the various fiduciary categories, but rather at the possible working relationship between the various kinds of corrective rules and responses that are united under the rubric of a singular fiduciary regulatory framework. It is, in other words, imperative to distinguish between different kinds or modes of fiduciary breach, and to match the content or extent of a fiduciary's obligation with the actual structural characteristics of the particular physical arrangement under examination.[96] This is because the content or extent of a fiduciary's responsibilities, and the law's corrective response to any actual or possible breach of those responsibilities, are affected by, and hence must vary with, the particular powers and opportunities enjoyed by the fiduciary, or type of fiduciary, in question. Needless to say, being a 'fiduciary' does not subject one to every incident of the express trust, for it is unnecessary to impose any higher obligation on a fiduciary than is needed to maintain the integrity of the particular limited-access relation, be it status-based or fact-based in nature.

The process of matching content with factual structure holds important insights for understanding Class 2 'relational' undue influence as it existed before 1985 in the United Kingdom. In particular, it assists in locating that form of undue influence within the wider structure of fiduciary regulation,

[96] A process that Flannigan calls 'matching content with structure': Flannigan, 'The Fiduciary Obligation,' above n 47, at 320. The idea essentially refers to the assignment of accountability over the full range of the limited access in a particular arrangement, without going beyond that access. *Cf* also *Attorney-General v Blake* (1997) [1998] Ch 439 (CA) 454–5 (Lord Woolf MR).

while defensibly maintaining its discrete doctrinal administration separate from general fiduciary law. Although both vigilant trust and deferential trust are forms of 'fiduciary' trust by virtue of the fact that each leads to 'fiduciary access' and hence to a common toxic mischief that the law seeks to avoid, this is not to say that the legal scope of protection afforded to the respective trust types must be identical for all intents and purposes. Consider vigilant trust. Where vigilant trust has been reposed in a fiduciary, liability is unconditionally strict. This is because vigilant trust is capable of being breached in many and varied ways, all of which are generally unappreciated by the trusting party or difficult to detect. The law accordingly holds the vigilantly trusted party strictly liable for all actual or potential conflictual acts or states of affairs that fall within the scope of the trusting relation or function, no matter how minor or innocent such acts or states of affairs may be. Moreover, when the victim of a breach of vigilant trust seeks transaction avoidance (as opposed to, say, a personal accounting or compensatory remedy),[97] causation between the breach and the impugned transaction is irrelevant from the standpoint of the victim's entitlement to rescind. That is because unauthorised entry into a transaction that is inconsistent with the fiduciary's duty of loyalty is itself a breach of duty 'patent at the creation of the very thing which is to be set aside'.[98]

Turning to deferential trust, or limited-access relations of influence, a different legal approach is evident. The focus is now on the special influential capacity enjoyed by the ascendant party in virtue of the subordinate party's known deference toward the former in relation to matters falling within the area of his or her authority, expertise or responsibility. In contrast to the vigilant trust situation, therefore, there is only one possible form of objectionable conduct that can occur: the effective and conflictual exercise by the fiduciary of that influence over the subordinate party for the fiduciary's own ends, which, at least pre-*Morgan*, is what defined all pure acts of relational undue influence. When such a use of influence causes the impugned transaction, the consent brought to that transaction by the subordinate party cannot be treated as a expression of that party's 'free and informed' will, so as to render him or her fully legally responsible for that consent, as his or her will was compromised by the unauthorised conflictual use of the ascendant party's special influence.

[97] It has been argued that causation should also not be required when an accounting remedy is sought after breach of the no-profit rule: see V Vann, 'Causation and Breach of Fiduciary Duty' [2006] *Singapore Journal of Legal Studies* 86. *Cf Regal (Hastings) Ltd v Gulliver* [1967] 2 AC 134 (HL) 144; *Gray v New Augarita Porcupine Mines Ltd* [1952] 3 DLR 1 (PC) 15. Views have been expressed to the opposite effect: see C Mitchell, 'Causation, Remoteness, and Fiduciary Gains' (2006) 17 *Kings College Law Journal* 325; *Chirnside v Fay*, above n 53 (concerning a joint venture).

[98] *Maguire v Makaronis* (1997) 188 CLR 449 (HCA) 467.

Causation now becomes relevant because, unlike in the vigilant trust situation, limited-access influence that results from deferential trust can only be effectuated instrumentally through the deferentially trusting party, who must thus be involved at the critical decision-making (but not necessarily the implementational) stage—that is, at the time of transacting. It may be that, despite the defendant's apparent or possible use of special influence over the plaintiff to effectuate the impugned transaction, the latter actually made the transactional decision, independently of the former's influence. In other words, the plaintiff may have in fact acted independently, either because the defendant did not actually exercise influence over the plaintiff or, if he or she did attempt to do so, this influence was ineffectual in bringing about the impugned transaction. Fiduciary liability may also be less strict in the deferential trust situation than in the vigilant trust cases, as even if the defendant had in fact enjoyed fiduciary influence over the plaintiff that was effectual in producing the impugned transaction, some formulations of the relational undue influence doctrine suggest that the defendant may still be absolved of liability (and thus retain any value received from the plaintiff) by showing that no unfair procedural or strategic advantage had been taken of that influence.[99] That might occur, for instance, because the defendant was, at the time of the transaction, justifiably ignorant of the plaintiff's deferential trust toward the defendant, or was entitled reasonably to believe that the plaintiff had been independently and competently advised on both the conflict and the general wisdom of what he or she was about to do. In other words, any conflict in the relational undue influence context is managed through a 'fair-dealing' rule, as an application of the overarching no-conflict rule, rather than a pure 'self-dealing' rule.[100]

Yet for all that, care must be taken not to overstate the contrast, since the defence to an alleged breach of trust, whether vigilant or deferential,

[99] I should emphasise that 'procedural fairness' must be shown, and that merely showing 'substantive fairness' does not alone suffice to avoid liability in this context (although substantive fairness may bolster a defence of procedural fairness). To be sure, giving too much weight to the existence of substantive fairness in the fiduciary context raises the usual detection and evidentiary concerns, in that it encourages the possibility of cosmetically manipulated transactions.

[100] Note that there is a parallel type of analysis in relation to the 'self-dealing' and 'fair-dealing' rules where trustees purchase trust property or a beneficiary's beneficial interest. Despite Megarry V-C's suggestion to the contrary in *Tito v Waddell (No 2)* [1977] Ch 106, 225, 241, both of those rules should now be seen as contextual regulation of the generic no-conflict rule: see, eg, E Simpson, 'Conflicts' in Birks, above n 56, at ch 3, especially at 82ff; M Conaglen, 'A Re-Appraisal of the Fiduciary Self-Dealing and Fair-Dealing Rules' [2006] *CLJ* 366. Note also that although relational undue influence does not involve 'self-dealing' in its purest sense (that is, where the selling/purchasing fiduciary controls the subject matter on both sides of the transaction), self-dealing nevertheless does occur in the sense and to the extent that the fiduciary's influence operates to control both sides of the transaction. *Cf* Shepherd, above n 18, at 156.

remains identical in both situations, namely, that the plaintiff 'fully consented' to the benefit received, which consent must include the plaintiff having received disclosure of the conflict that would otherwise exist. How the different types of trust operate, however, determines the different steps that the defendant must take toward securing the plaintiff's consent in that regard. Only undue influence requires the act of emancipation that, in a practical sense, operates to eliminate the conflictual state of affairs, as well as securing the plaintiff's personal transactional consent to the impugned benefit. This should not, however, be understood as implying that deferential trust is somehow less worthy of protection than vigilant trust, or that the legal scope of protection is somehow narrower. The differences between the respective forms of fiduciary regulation and specific legal responses are simply consequences of the varying physical phenomena that attend the parties' relationship or arrangement in each situation: the nature of the different types of limited access (or fiduciary trust) involved, the manner in which each type of access or trust tends to operate, the specific types of opportunity that each tends to present to the fiduciary for diverting value to himself or herself (or to third parties), and the specific and varying forms of wrongdoing (opportunism) that might relevantly flow from such opportunities.

III. REFLECTIONS ON THE DISINTEGRATION OF RELATIONAL UNDUE INFLUENCE AS A FIDUCIARY RULE—FROM *MORGAN* TO *ETRIDGE* IN THE HOUSE OF LORDS

A. *National Westminster Bank plc v Morgan*

As mentioned earlier, the essential fiduciary character of relational undue influence in the United Kingdom was suppressed or diminished in 1985 with the unanimous decision of the House of Lords in *Morgan*. There the Law Lords expressly rejected the so-called 'public policy' approach to relational undue influence. Lord Scarman, with whom the other Law Lords agreed, held that the basis for relief in this area was 'not a vague "public policy" but specifically the victimisation of one party by the other'.[101] On this approach, the party seeking exculpation from the transaction must, in addition to proving a special relationship of influence *inter partes*, show that the resultant bargain was 'manifestly disadvantageous' to him or her,

[101] Above n 30, at 705, citing Lindley LJ in *Allcard v Skinner*, above n 7, at 182–3. Interestingly, in *Etridge* the House of Lords did not comment directly on the compatibility (or otherwise) of the two approaches. However, the language employed in the various judgments is entirely consistent with that of the 'victimisation' approach and the 'public policy' approach is not discussed at any point.

as judged by the (obviously) substantively unfair nature of the impugned transaction itself. No presumption of undue influence could therefore arise in respect of a transaction that provided 'reasonably equal benefits for both parties'.[102] His Lordship also preferred to 'avoid the term "confidentiality" as a description of the relationship which has to be proved' in the presumptive (Class 2) cases.[103]

Lord Scarman's judgment, however, is problematic in many respects. Although some of his errors were identified and corrected in later House of Lords decisions—of which more shortly—the effects of the dismissal of the 'public policy' (fiduciary) approach to relational undue influence in *Morgan* are still evident in the subsequent enunciation of the 'first principles' of undue influence in *Etridge*. This is despite the fact that, amid those two decisions, Lord Browne-Wilkinson, in *Pitt*,[104] flagged the possible need to consider 'the exact limits of *Morgan* … in the future'. In particular his Lordship said:

> The difficulty is to establish the relationship between the law [of undue influence] laid down in *Morgan* and the long established principle laid down in the abuse of confidence cases viz. the law requires those in a fiduciary position who enter into transactions with those to whom they owe fiduciary duties to establish affirmatively that the transaction is a fair one … [This] principle is founded on considerations of general public policy, viz. that in order to protect those to whom fiduciaries owe duties *as a class* from exploitation by fiduciaries *as a class*, the law imposes a heavy duty on fiduciaries to show the righteousness of the transactions they enter into with those to whom they owe such duties. … Unfortunately, the attention of this House in *Morgan* was not drawn to the abuse of confidence cases and therefore the interaction between the two principles (if indeed they are two separate principles) remains obscure[105]

As we shall see, the House of Lords subsequently did not, in *Etridge*, confront this 'difficulty' or attempt directly to resolve the continuing 'obscurity' between the two principles.

Turning now to Lord Scarman's reasoning in *Morgan*, and in particular his statement that the principle justifying interference with a transaction for undue influence is 'victimisation' rather than 'vague "public policy,"' it must first be noted that his Lordship relied for support on a series of passages from Lindley LJ's judgment in *Allcard v Skinner*.[106] In one passage Lindley LJ opined that the undue influence doctrine is founded on the principle that 'it is right and expedient to save [people] from being victimised by other people', rather than on the ground that they should

[102] *Morgan*, above n 30, at 704.
[103] *Ibid*, at 709.
[104] Above n 7, at 209.
[105] *Ibid*, at 209 (emphasis in original).
[106] Above n 7.

simply be saved 'from the consequences of their own folly'.[107] Pausing here for a moment, it is obvious that this passage does not alone necessitate rejection of the 'public policy' approach to relational undue influence, for that approach is itself concerned entirely with regulating against the risk of victimisation by influential fiduciaries in circumstances where proof of actual victimisation by way of relational undue influence is typically difficult if not impossible. Although it might produce a quite artificial finding of victimisation in a particular case (when in fact the influential party may have been factually innocent), in no way is the public policy approach motivated by any desire to save parties from bad bargains or regretted gifts. On the contrary, as acknowledged in the above passage from Lord Browne-Wilkinson's judgment in *Pitt*, the public policy approach in fiduciary contexts exists precisely 'in order to protect those to whom fiduciaries owe duties *as a class* from exploitation by fiduciaries *as a class*'. The 'public policy' and 'victimisation' approaches are not, in fact, mutually exclusive of each other.[108] Public policy merely justifies and is expressed through the distinctive process of reasoning from the duty (the limited-access relation of influence) to the breach (abuse of that relation or victimisation in fact), that is, through a presumption of undue influence based on the mere existence of a transaction *inter se*, together with the fact that 'the character of the relation itself is never enough to explain the transaction and to account for it without suspicion of confidence abused'.[109]

The objection to Lord Scarman's use of Lindley LJ's distinction between victimisation and mere regretted transactions aside, once making the move to factual victimisation for additional reasons,[110] his Lordship nonetheless proceeded to secure his error in rejecting the public policy approach to relational undue influence by making it clear in his judgment that he viewed the wrong of 'victimisation' as residing, at least in part, in the substance of the impugned transaction itself. I say 'at least in part' here because Lord Scarman's judgment is actually unstable on this point. It vacillates unsteadily, and seemingly without awareness, between, on the one hand, viewing substantive unfairness (or 'manifest disadvantage') merely as part of the factual matrix that would leave 'room for the court to presume that [the impugned transaction] resulted from the exercise of

[107] *Ibid*, at 182–3, cited in *Morgan*, above n 30, at 705.

[108] *Cf* Sheridan, above n 12, at 87: 'fraud and things of confidence do not constitute a mutually exclusive dichotomy'.

[109] *Yerkey v Jones* (1939) 63 CLR 649 (HCA) 675 (Dixon J).

[110] That is to say, in addition to relying on Lindley LJ's remarks in *Allcard v Skinner*, Lord Scarman relied on two Privy Council decisions—*Bank of Montreal v Stuart* [1911] AC 120 (HL) 137 and *Poosathurai v Kannappa Chettiar* (1919) LR 47 IA 1 (PC) 3—which he considered supported his particular victimisation approach: see *Morgan*, above n 30, at 706–707.

undue influence'[111]—that is, it would allow the court to infer victimisation in fact, or generate a 'presumption' of undue influence—and, on the other hand, regarding it as an ingredient of the wrong of 'undue influence' itself. In Lord Scarman's words, there is a:

> need to show that the transaction is wrongful in the sense explained by Lindley LJ [in *Allcard v Skinner*—namely, that 'the gift is so large as not to be reasonably accounted for on the ground of friendship, relationship, charity, or other ordinary motives on which ordinary men act'[112]—] before the court will set aside a transaction whether relying on evidence or the presumption of the exercise of undue influence.[113]

Although these words were later to steer lower courts into error, in particular by causing them to require proof of manifest disadvantage in relation to Class 1 non-relational undue influence claims,[114] the House of Lords promptly took corrective action in *Pitt*, pointing out that Lord Scarman's speech 'was primarily concerned to establish that disadvantage had to be shown, not as a constituent element of the cause of action for undue influence, but in order to raise a presumption of undue influence [within] Class 2'.[115] Lord Scott of Foscote also noted in *Etridge* the obvious circularity in Lord Scarman's reasoning that no 'presumption of undue influence [could] arise from the evidence of the relationship of the parties without also evidence that the transaction itself was wrongful in that it constituted an advantage taken of the person subjected to the influence which, failing proof to the contrary, was explicable only on the basis that undue influence had been exercised to procure it'.[116] How could a transaction be adjudged 'wrongful' before it was proven to have been procured by undue influence?

Now, regardless of what it is that triggers the presumption of undue influence in the Class 2 cases,[117] it is clear that Lord Scarman did not formulate his manifest disadvantage test by reference to the fiduciary rationale that had, until his own rejection of the 'public policy' approach, governed the operation of the presumption according to the substantive policy that justified the creation and application of the presumption in the first place. Lord Scarman did not, after all, consider the influential relation between the parties to a Class 2 undue influence claim to be 'fiduciary' in nature at all. Commenting adversely on the Court of Appeal's factual

[111] *Morgan*, above n 30, at 707.
[112] Above, n 7, at 185.
[113] *Morgan*, above n 30, at 706.
[114] See, eg, *Bank of Credit and Commerce International SA v Aboody* (1988) [1990] 1 QB 923 (CA) 967; *Contractors Bonding Ltd v Snee* [1992] 2 NZLR 157 (CA).
[115] Above n 7, at 208.
[116] Above n 6, at [155].
[117] I have discussed this elsewhere in Bigwood, above n 40, at 423–56.

classification of the relationship between the bank and Mrs Morgan as being one of 'confidence in which [Mrs Morgan] was relying on the bank manager's advice', and that 'the confidentiality of the relationship was such as to impose upon [the manager] a "fiduciary duty of care,"'[118] Lord Scarman said:

> I believe that the Lords Justices were led into a misinterpretation of the facts by their use, as is all too frequent in this branch of the law, of words and phrases such as 'confidence,' 'confidentiality,' 'fiduciary duty.' There are plenty of confidential relationships which do not give rise to the presumption of undue influence (a notable example is that of husband and wife …); and there are plenty of non-confidential relationships in which one person relies upon the advice of another, eg many contracts for the sale of goods.[119]

He then concluded his judgment by declaring a preference for avoidance of the term 'confidentiality' as a description of the relationship that must be proved in the presumptive cases;[120] the search, rather, was for a 'dominating influence of one over another,' which search required a 'meticulous examination of the facts'.[121]

The problem with Lord Scarman's criticism of the Court of Appeal's use of 'fiduciary' terminology in connection with the influential relation that must be shown in the presumptive cases is that it simply begs the question as to where the real points of similarity and difference lie between relational undue influence and wider fiduciary law. His reference to 'reliance upon the advice of another', for example, is simply unhelpful unless we first agree that we are only talking about reliance in the relevant sense, as importing something that leads to 'limited access to assets'.[122] 'Reliance' is open to expansive meanings and is thus too unstable as a descriptor for identifying a fiduciary obligation. And even assuming that the relation of husband and wife can be described as a 'confidential' one (again, in the relevant sense),[123] the reason why the presumption of undue influence does not apply to transactions between spouses relates not necessarily to the nature of the relationship involved but rather to the lack of natural suspicion surrounding generous dispositions *inter se*. In other

[118] *Morgan*, above n 30, at 702.
[119] *Ibid*, at 703.
[120] *Ibid*, at 709.
[121] *Ibid*.
[122] Taking Lord Scarman's 'sale of goods' example, for instance, although it is clear that I might rely on a seller's advice (eg that the goods I wish to acquire will suit my purposes or perform in a certain way), the seller's access to my assets in exchange for those goods is 'open' or 'unlimited'. In other words, I do not, or am not entitled to, rely on the seller to suspend self-interest in connection with the relationship or encounter in question, as its access to my money in return for the goods is for its own account and not mine.
[123] That is, as importing 'some quality beyond that inherent in the confidence that can well exist between trustworthy persons who in business affairs deal with each other at arm's length': *Lloyds Bank Ltd v Bundy*, above n 44, at 341 (Sir Eric Sachs).

words, partiality is expected between spouses (and similarly intimate parties); there is ordinarily nothing that prima facie 'calls for explanation' so as to trigger the presumption of undue influence.[124]

Indeed, the sort of question-begging evident in Lord Scarman's judgment seems to contaminate all denials of the essential 'fiduciary' nature of relational undue influence, whether they are absolute in nature[125] or qualified in some way (typically by the rider that the two bodies of doctrine 'overlap but do not coincide').[126] For myself, I find all attempts at differentiation (so far) to be unconvincing, at least when they are directed at the conceptual level. The points of distinction are either asserted or purely formal in nature. One cannot ignore the ubiquity of fiduciary obligation in our society, as Fletcher Moulton LJ's famous 'errand boy' example in *Re Coomber*[127] made clear,[128] and in particular the fact that

[124] *Bank of Montreal v Stuart*, above n 110; *Yerkey v Jones*, above n 109; *Howes v Bishop and Wife* [1909] 2 KB 390 (CA) 396 (Lord Alverstone), 400–403 (Farwell J); *Etridge*, above n 6, at [19] (Lord Nicholls). Additional phenomena may attend the husband and wife relationship so as to make it a Class 2(B) situation: *Etridge*, above n 6, at [130] (Lord Hobhouse), [283] (Lord Scott, on *Barclays Bank plc v Coleman* [2001] QB 20 (CA) in which the husband and wife were members of the Hasidic Jewish community, which factually generated a relationship of complete trust and confidence between a wife and her husband in relation to financial matters).

[125] See, eg, WHD Winder, 'Undue Influence and Fiduciary Relationship' (1940) 4 *The Conveyancer and Property Lawyer* 274, to the effect that the law of undue influence does not apply to transactions for value, only to gifts—abuse of confidence is the proper ground for interference with transactions for value; G Spencer Bower, *The Law Relating to Actionable Non-Disclosure* (London, Butterworths, 1915) 272–3, explaining that the two relations co-exist and coalesce, but the duties that arise out of them are merely 'accidental' and must be distinguished; *cf* G Spencer Bower, AK Turner and RJ Sutton, *The Law Relating to Actionable Non-Disclosure*, 2nd edn (London, Butterworths, 1990) §§ 16.07, 21.04–21.06, agreeing with the statements in first edition of the work; RP Austin, 'The Corporate Fiduciary: Standard Investments Ltd v Canadian Imperial Bank of Commerce' (1986) 12 *Canadian Business Law Journal* 96, 97: 'the two doctrines are conceptually different'; P Birks and NY Chin, 'On the Nature of Undue Influence' in J Beatson and D Friedmann (eds), *Good Faith and Fault in Contract Law* (Oxford, Oxford University Press, 1995) 91: 'It is less useful, even dangerous, to create a close relationship between undue influence and breach of fiduciary duty'; S Worthington, above n 50, at 503, fn 18, arguing that the law of undue influence 'is completely independent of and distinct from' fiduciary law.

[126] See, eg, PV Baker and P St J Langan, *Snell's Principles of Equity*, 28th edn (London, Sweet & Maxwell, 1982) 544, cited with approval by Slade LJ in *Bank of Credit and Commerce International SA v Aboody*, above n 114, at 962 (although the House of Lords in *Pitt*, above n 7, at 208, comments that the Court of Appeal in *Aboody* regarded the abuse of confidence (breach of fiduciary duty) line of cases as 'a wholly separate doctrine of equity'); M Cope, *Duress, Undue Influence and Unconscientious Bargains* (Sydney, Law Book Co, 1985) 79ff; RP Meagher, JD Heydon and MJ Leeming, *Equity: Doctrines and Remedies*, 4th edn (Sydney, Butterworths, 2002) 514.

[127] [1911] 1 Ch 723 (CA) 728: 'Fiduciary relations are of many types; they extend from the relation of myself to an errand boy who is bound to bring me back my change up to the most intimate and confidential relations which can possibly exist between one party and another where the one is wholly in the hands of the other because of his infinite trust in him.'

[128] Flannigan, 'Commercial Fiduciary Obligation,' above n 49, at 911, eg, complains that we often lose sight of the ubiquitous and generic nature of fiduciary obligation, which is a fact that should always inform our approach to fiduciary issues: 'We lose track of this [fact], it

fiduciary tasks or functions routinely arise within relationships that are not 'fiduciary' for all intents and purposes.[129] This observation underscores the purely formal nature of the common assertion that the 'fiduciary relationship' and the 'relationship of influence' (for the purposes of a Class 2 undue influence claim) overlap but are not identical. Consider, for instance, the following textbook example, which is very commonly encountered:

> some relationships of influence are clearly not also fiduciary in character (for example, parent and child) and it is clear that the categories are not co-incident though they have limited areas in common.[130]

But tasks or functions that happen to be performed by actors who are not typically fiduciaries for all intents and purposes must nevertheless be regarded as 'fiduciary' to the extent that they involve the actor acquiring direct or indirect access to another's assets for the defined and limited purpose of the parties' relationship or undertaking, rather than for the actor's own, unencumbered purposes.[131] So, for example, although the parent–child and guardian–ward relationships are not formally regarded as 'fiduciary relationships' *in toto*, parents and guardians are nevertheless 'fiduciaries' in virtue of, and to the extent of, the 'tutelary' and 'advisory' functions they each perform in respect of their respective unemancipated children and wards.[132] Such functions give the functionary a peculiar ability to influence the choices and actions of the beneficiary who, in the context of the specific physical arrangement between the parties, is dependent on the functionary performing his or her special tasks in a conscientious and disinterested manner. In other words, the special tasks become the source of the deferential trust (the limited-access relation of

seems, when we ask whether a particular *relationship* is fiduciary. It is preferable to ask whether there are fiduciary *obligations* arising from a specific physical arrangement. A mechanic is not generally considered to be in a fiduciary relationship with customers. Yet a mechanic will have, along with certain contract and tort obligations, certain fiduciary *obligations*. In that sense (and to that limited extent), the relationship is fiduciary' (emphasis in original). See also Flannigan, 'Fiduciary Mechanics', above n 67.

[129] Hence the label 'fiduciary' is inadequate per se to distinguish between those (fiduciary) relationships that will give rise to a presumption of undue influence and those that will not. Not all relationships that are 'fiduciary' also involve 'influence' of the kind required to attract relational undue influence regulation. This point is well recognised in the case law: *Re Coomber*, above n 127, at 726–7 (Cozens-Hardy MR), 730 (Buckley LJ); *Tufton v Sperni*, above n 44, at 530 (Jenkins LJ); *Cowen v Piggott* (1989) 1 Qd R 41 (QSC) 44 (McPherson J, as quoted by Connelly J in the Full Court).

[130] Meagher, Heydon and Leeming, above n 126, at 514.

[131] Flannigan, 'Fiduciary Regulation of Sexual Exploitation,' above n 49, at 307. See also BH McPherson, 'Fiduciaries: Who Are They?' (1998) 72 *Australian Law Journal* 288, 288.

[132] Likewise of the doctor–patient relationship: see *Breen v Williams*, above n 55, at 83 (Brennan CJ), 92 (Dawson and Toohey JJ), 107–108 (Gaudron and McHugh JJ), 134–5 (Gummow J). Patients repose trust in their doctors, who in turn owe confidentiality obligations and (are presumed to) acquire ascendancy over their patients for the purpose of undue influence regulation, but they are not otherwise fiduciaries.

influence) that attracts a 'fiduciary' obligation for the purposes of Class 2 relational undue influence law. Equally, guardians, doctors, spiritual advisers and the like, are 'fiduciaries' by virtue and to the extent of their ability to influence their wards, patients and penitents (respectively) in their capacity as 'impartial advisers' to such persons. Indeed, it has long been recognised that relationships can have varied purposes, some of which, or points within which, attract a 'fiduciary' obligation while others do not.[133] Relational undue influence claims, which imply breach of deferential trust, are cardinal illustrations of this principle.

B. *Royal Bank of Scotland plc v Etridge (No 2)*

Morgan was not well received by all commentators and courts,[134] at least in respect of some of its points of law. In *Goldsworthy v Brickell*,[135] for example, Nourse LJ refused to accept that, regarding the type of relation that attracted the presumption of undue influence, the House of Lords 'could have intended, sub silentio, to overrule not only *Tufton v Sperni* … but many other leading cases from *Huguenin v Baseley* … onwards'.[136] In *Etridge*, though, the House of Lords could have rescued the fiduciary account of relational undue influence but did not. Although one might have hoped that it would have directly and robustly examined the interaction, if any, between relational undue influence and fiduciary obligation—as had, after all, been earlier signalled as a possibility by Lord Browne-Wilkinson in *Pitt*—that simply did not occur.[137] It seems clear that the House of Lords was not interested in reversing the trend set by *Morgan*, but the justification for that stance remains far from clear. There is in *Etridge* virtually no critical dissection of the earlier

[133] *New Zealand Netherlands Society 'Oranje' Inc v Kuys* [1973] 2 All ER 1222 (PC) 1225 (Lord Wilberforce); *Hospital Products Ltd v United States Surgical Corporation*, above n 62, at 98 (Mason J); *Noranda Australia Ltd v Lachlan Resources NL* (1988) 14 NSWLR 1 (SC) 15 (Bryson J): 'a person under a fiduciary obligation to another should be under that obligation in relation to a defined area of conduct, and exempt from the obligation in all other respects. Except in the defined area, a person under a fiduciary duty retains his own economic liberty'. This parallels closely courts' treatment of the purchasing rule in fiduciary law. Thus, as Lord Blackburn observed in *McPherson v Watt* (1877) 3 App Cas 254 (HL) 270–71, eg, a solicitor–client case, 'If [the fiduciary] purchases … in a matter totally unconnected with what he was employed in before, no doubt an attorney may purchase from one who has been his client, just as any stranger may do … being in no respect bound to do more than any other purchaser would do. But when he is purchasing from a person *property with respect to which the confidential relation has existed or exists*, it becomes wrong of him to purchase without doing a great deal more than would be expected from a stranger' (emphasis added).

[134] A convenient summary of the various academic views on *Morgan*, offered shortly after the decision, can be found in *Geffen v Goodman Estate*, above n 16, at 222–6 (Wilson J).

[135] Above n 44.

[136] *Ibid*, at 406.

[137] Only Lord Clyde mentions Lord Browne-Wilkinson's passage, but he does not respond to its invitation: *Etridge*, above n 6, at [104].

authorities on the subject, and much of the elaboration of the 'first principles' of undue influence reads as assertion rather than analysis. The signals sent about the nature of liability for relational undue influence are mixed, and there is appreciable internal inconsistency in the individual judgments.[138] As a result, commentators on the case seem to have taken slightly different messages from it, as have later courts purporting to describe and apply the now reigning law.[139]

By way of illustration of the doctrinal and conceptual instability created by the exposition of undue influence in *Etridge*, consider first Lord Nicholls' description of that form of 'unacceptable conduct' that 'arises out of a relationship between two persons where one has acquired over another a measure of influence, or ascendancy, of which the ascendant person then takes unfair advantage'[140]—cases involving what we have been calling 'relational' or 'Class 2' undue influence. He distinguishes those cases from another form of unacceptable conduct (in the manner of wrongful persuasion) that equity broadly identified as falling within the principles of undue influence, namely, 'overt acts of improper pressure or coercion such as unlawful threats'[141]—cases involving so-called 'non-relational' or 'Class 1' undue influence. In describing the nature of the influence that is abused in the 'relationship' cases, Lord Nicholls explains how it 'provides scope for misuse without any specific overt acts of persuasion'[142] (hence its very nature makes monitoring and detection difficult). He continues, in language that is classically 'fiduciary' in its purport, as demonstrated in particular by the following emphasised words:

> The relationship between two individuals may be such that, without more, one of them is disposed to agree a course of action proposed by the other. Typically this occurs when one person *places trust in another to look after his affairs and interests*, and the latter *betrays this trust by preferring his own interests*. He abuses the influence he has acquired. In *Allcard v Skinner* … Lindley LJ … described this class of cases as those in which it was the *duty of one party to advise the other or to manage his property for him*. In *Zamet v Hyman* … Lord Evershed MR referred to relationships where one party owed the other an obligation of candour and protection.[143]

[138] Although all of the Law Lords purport to agree with Lord Nicholls' judgment.

[139] See, eg, *Hammond v Osborn* [2002] EWCA Civ 885; *Glanville v Glanville* [2002] EWHC 1587 (Ch); *Niersmans v Presticcio* [2004] EWCA Civ 372.

[140] *Etridge*, above n 6, at [8].

[141] *Ibid*. The reference here to 'unlawful threats' is immediately anomalous, as such threats, if effectual, would clearly constitute duress at common law, whereas equity tended to regulate coercive conduct that fell below strict illegality such as lawful but unconscientious pressure. Lord Hobhouse also stated that Class 1 undue influence 'is capable of including conduct which might give a defence at law, for example, duress and misrepresentation': *ibid*, at [103].

[142] *Ibid*, at [9].

[143] *Ibid* (emphasis added).

Lord Nicholls mentions 'parent and child' as an example of the type of relation in equitable contemplation here, but he rightly accepts that the relationships in which the Class 2 principles fall to be applied are incapable of being listed exhaustively.[144]

So far so good, at least for the fiduciary account, but in the very next paragraph of his judgment Lord Nicholls potentially explodes the fiduciary boundaries of the so-called 'relationship' cases:

> Even this test [of whether 'one party has reposed sufficient trust and confidence in the other'] is not comprehensive. The principle [applicable to the 'relationship' cases] is not confined to cases of abuse of trust and confidence. It also includes, for instance, cases where a vulnerable person has been exploited.[145]

He then asserts that 'no single touchstone [exists] … for determining whether the principle is applicable.' He observes:

> Several expressions have been used in an endeavour to encapsulate the essence: trust and confidence, reliance, dependence or vulnerability on the one hand and ascendancy, domination or control on the other. None of these descriptions is perfect. None is all embracing. Each has its proper place.[146]

Here, there are problems with Lord Nicholls' beliefs and analysis. First, by appearing to extend the scope of the relational undue influence principle beyond limited-access relations of influence (deferential trust) to wider, arm's-length power–vulnerability relations (exploitable disadvantage), he fails to observe the distinctive character of relational undue influence circumscribed by the fiduciary principle, and indeed risks rendering Class 2 undue influence conceptually and practically indistinguishable from mere unconscionable dealing, particularly when the exploitation involved is of the 'passive' kind, or by omission only.[147] And although it is not clear that it was Lord Nicholls' intention to collapse relational undue influence into unconscionable dealing, for there is no explicit attempt anywhere in *Etridge* to reconcile undue influence with its equitable sibling, it is nonetheless a reasonable reading of the words that he chose to express the

[144] *Ibid*, at [10].
[145] *Ibid*, at [11].
[146] *Ibid*.
[147] Not everyone considers such conceptual intermixture to be undesirable. See, eg, I Hardingham, 'Unconscionable Dealing' in P Finn (ed), *Essays in Equity* (Sydney, Law Book Co, 1985) 18; A Phang, 'Undue Influence—Methodology, Sources and Linkages' [1995] *Journal of Business Law* 552; D Capper, 'Undue Influence and Unconscionability: A Rationalisation' (1998) 114 *LQR* 479; A Phang and H Tjio, 'The Uncertain Boundaries of Undue Influence' [2002] *Lloyds Maritime and Commercial Law Quarterly* 231, 232–4, 241–3; A Phang and H Tjio, 'Drawing Lines in the Sand? Duress, Undue Influence and Unconscionability Revisited' (2003) 11 *Restitution Law Review* 110, 117ff. *Cf* also J Devenney and A Chandler, 'Unconscionability and the Taxonomy of Undue Influence' [2007] *Journal of Business Law* 541.

jurisdiction and the natural effect of the approach to the burden and methods of proof in connection with undue influence claims in general—of which more shortly.

Moreover, the extension of relational undue influence to encompass wider power–vulnerability situations is simply asserted in Lord Nicholls' judgment. No supporting authority is mentioned and no justification is given. One is left instead to speculate. Certainly no example of relational undue influence occurring outside a (subjectively?) 'trusting and confidential' relationship is given at the relevant point in his judgment. It is likely that he merely had in mind, and so was affirming *sub silentio*, a point that had been made in the Court of Appeal below, where Stuart-Smith LJ suggested that undue influence was not confined to 'abuse of confidence' principles.[148] But neither of the examples that Stuart-Smith LJ cited as an example of 'undue influence' by exploitation of a person who is vulnerable otherwise than by reason of being in a trusting and confidential relationship with her exploiter—*Allcard v Skinner* and *Re Craig*—are inconsistent with the traditional fiduciary account of the Class 2 cases.[149] As seen in *Johnson v Buttress*, for example, a limited-access relationship of influence can equally result from known 'ascendancy and dependence' *inter partes*, if that ascendancy and dependence is of a 'deferential' kind, as from an actual concession of (deferential) trust and confidence in another. Indeed, earlier in his judgment Lord Nicholls himself cited *Allcard v Skinner* as a typical 'trust'-type case,[150] which is at variance with his apparent acceptance of Stuart-Smith LJ's observations in the Court of Appeal below. At least on one reading of *Allcard v Skinner*—and granted the judgments there are open to diverse interpretations[151]—the lady superior in that case enjoyed 'limited access' to the claimant's assets, and hence was subject to fiduciary accountability, by virtue of the special 'advisory' function that is habitually performed by spiritual leaders in relation to their devotees (not

[148] [1998] 4 All ER 705, [8]: 'The equitable doctrine of undue influence … is not confined to cases of abuse of trust and confidence; it is also concerned to protect the vulnerable from exploitation. It is brought into play whenever one party has acted unconscionably in exploiting the power to direct the conduct of another which is derived from the relationship between them. This need not be a relationship of trust and confidence; it may be a relationship of ascendancy and dependence.' Again, and with respect, this passage misses the point that it is 'limited access' regardless of its source (trust and confidence, ascendancy and dependence, or whatever) that attracts the Class 2 jurisdiction. The remainder of the paragraph in Stuart-Smith LJ's judgment, however, confirms that he was talking of Class 1 undue influence occurring within the context of 'some close and confidential relation to the donor,' as he begins the sentence following the above-quoted words with: 'In such cases *actual* undue influence has been said to involve' (emphasis added).

[149] Above n 7 and above n 6 respectively.

[150] *Etridge*, above n 6, at [9].

[151] As Charlotte Smith's recent essay shows: C Smith, '*Allcard v Skinner* (1887)' in C Mitchell and P Mitchell (eds), *Landmark Cases in the Law of Restitution* (Oxford, Hart Publishing, 2006) ch 8.

to mention by virtue of the particular rules of the sisterhood by which the claimant donor had become bound). And in *Re Craig*, the companion/housekeeper became an ad hoc 'fiduciary' because of the elderly donor's known dependency on her 'for his comforts and emotionally for her companionship and for her participation in his business affairs', and because of her comparative powers, abilities, competency, managing disposition, strong personality and physical and mental toughness.[152] As a result of these features of the parties' respective positions, Ungoed-Thomas J concluded that the companion/housekeeper 'had a duty in the circumstances to advise [the elderly donor], and if not to advise him at any rate to take care of him, in the management and disposal of his property'.[153] This, certainly, much like *Johnson v Buttress* before it, is consistent with a 'fiduciary' function being imposed on the companion/housekeeper by virtue of a limited-access relation of influence having resulted from the cumulative circumstances of the case, in particular because a fact-based deferentially trusting relationship had been established.

In the final analysis, Lord Nicholls' suggestion that 'no single touchstone [exists] for determining whether the principle [in the 'relationship cases'] is applicable' is telling of the House of Lords' failure to appreciate fully the fiduciary nature of all relational undue influence claims, and why that is so. There is, in fact, a single determinative yardstick for operation of the principle in the Class 2 cases—namely, whether the defendant's access to the plaintiff's assets is qualified by a limited purpose—and none of the expressions mentioned by Lord Nicholls that have been used to capture the essence of the cases—'trust and confidence, reliance, dependence or vulnerability on the one hand and ascendancy, domination or control on the other'—exist meaningfully or authoritatively in isolation of that inquiry. None of them alone expresses the essence of the requisite relation of influence; all are merely proxies for 'limited access to assets' or else descriptions of various factual phenomena that might, severally or in combination, with or without the aid of presumption,[154] lead a court to determine that a limited-access relation of influence existed between the parties.[155] The imposition of fiduciary obligation and regulation simply follows from that ultimate finding. Ironically, although Lord Nicholls states that each expression 'has its proper place,' nowhere are we told what that is exactly. The answer, though, is that each expression is useful because each, properly understood, performs an important signalling function as to the possible existence of a limited-access relation of

[152] Above n 6, at 119–20 (Ungoed-Thomas J).
[153] *Ibid*, at 120.
[154] That is, in the Class 2(A) cases, where a presumption of limited-access is utilised.
[155] Which is really why, in Lord Nicholls' words, none of the expressions is 'perfect' or 'all embracing'.

influence *inter se*. That type of relation is in fact the 'single touchstone' that the Law Lords in *Etridge* could have identified based on the conventional jurisprudence available to it before *Morgan*, but whose obscurity it managed to perpetuate instead.

The other Law Lords in *Etridge* send equally mixed signals as to the essential nature of the relationship that attracts the principles of relational undue influence. Lord Hobhouse, for example, describes the Class 2, 'presumed' undue influence cases as 'necessarily involv[ing] some legally recognised relationship between the two parties,' the result of which 'one party is treated as owing a special duty to deal fairly with the other. … *Typically they are fiduciary or closely analogous relationships*.'[156] His use of the word 'typically' would indicate that other types of relationship are envisioned here, but no clue is given as to what those are. Reference to the relationship being one in which one party owes a 'special duty to deal fairly with the other' is vague and begs the question entirely. General arm's-length relationships can engender duties of 'fairness' (for example not to coerce or mislead), as can specific relationships that are not part of relational undue influence law such as relationships involving special disadvantage that attract the potential operation of the unconscionable dealing doctrine. Lord Hobhouse cites an example of a solicitor owing a duty to 'deal fairly' with a client when purchasing property from him or her,[157] presumably within the scope of the solicitor's fiduciary function,[158] but then in turn fails to distinguish the regular breach of fiduciary duty situation from relational undue influence. The latter involves a breach of deferential trust only (and, depending on the facts, the solicitor may have been trusted in a vigilant sense only, or in a deferential sense only, or in both senses concurrently).[159] Lord Hobhouse also fails to acknowledge that the purchasing solicitor's burden of demonstrating 'fair dealing' in the example is not a manifestation of an obligation simply to 'act fairly' in relation to business transactions with a client. Rather, it is part and parcel of the singular and generic 'fiduciary' responsibility to act 'disinterestedly' toward him or her, in the sense that the fair-dealing rule that applies to fiduciaries who purchase from their obligees within the scope of the fiduciary relationship or function must be understood against the backdrop

[156] *Etridge*, above n 6, at [104] (emphasis added).

[157] *Ibid*.

[158] In *Allison v Clayhills* (1907) 97 LT 709 (Ch) 712, Parker J suggested that the presumption of undue influence would not apply if a solicitor bought a horse from a client who had retained him to conduct an action for slander, the relationship of influence existing only in respect of the latter.

[159] Clients tend to expect solicitors to act as their agents in implementing decisions (raising the intermediary cost mischief) and, in the same or another transaction, expect them also to advise in a disinterested manner (raising the undue influence mischief). Indeed, *Allcard v Skinner* can be seen as a case where the particular spiritual adviser was trusted in both the vigilant and the deferential senses: see especially above n 7, at 182 (Lindley LJ).

of an overarching 'no-conflict' rule that does not apply to purchasing parties who might otherwise be duty-bound to 'deal fairly' with their counterparty in the context of regular, arm's-length transactional encounters.

In his judgment Lord Scott speaks in terms of relationships of 'trust and confidence'[160] but appears to assign them no special meaning so as to indicate that the notion of fiduciary trust and confidence is necessarily intended thereby. For example, he mentions a wife having 'trust and confidence' in her husband's ability to make financial and business decisions, but of course this does not necessarily imply that 'limited access', hence 'fiduciary' obligation, is necessarily established *inter se*. Indeed, he distinguishes the relation of husband and wife from other relationships 'generally of a fiduciary character, where, *as a matter of policy*, the law requires the dominant party to justify the righteousness of the transaction'.[161] This suggests that Lord Scott might regard the 'presumption' in the Class 2(A) cases as something different than the 'presumption' in the Class 2(B) cases,[162] which he saw as 'doing no more than recognising that evidence of the relationship between the dominant and subservient parties, coupled with whatever other evidence is for the time being available, may be sufficient to justify a finding of undue influence on the balance of probabilities'.[163] Although this of course is to 'doubt the utility of the Class 2(B) classification',[164] presumably on redundancy grounds,[165] in the end analysis the point is asserted rather than validated in his judgment. Traditionally, Class 2(A) and Class 2(B) undue influence differed only in the manner of the claimant's proving the existence of the special 'influence' claimed to have been abused in the procurement or receipt of the impugned benefit, the existence of such influence being merely one ingredient in the relational undue influence claim, together with some sort of evidence as to its actual or possible abuse. The nature of the influence regulated in both cases, however, was otherwise identical, and it was limited-access, 'fiduciary'-type, influence.[166] The distinction between the two sub-classes of relational undue influence claim merely echoed the familiar distinction between 'status-based' and 'fact-based' fiduciaries that subsists in the wider fiduciary law. It followed, too, that the nature of the 'presumption' in each

[160] See, eg, *Etridge*, above n 6, at [158]–[159].
[161] *Ibid*, at [158] (emphasis added).
[162] This seems to have led some to assume that there are now three types of undue influence: actual, presumed and relational (a species of actual undue influence). See possibly Ferris, above n 47. However, the House of Lords makes it quite clear that, at least from a forensic standpoint, 'presumed' undue influence must also count as a form of 'actual' undue influence.
[163] Etridge, above n 6, at [161].
[164] *Ibid*.
[165] *Cf* Lord Hobhouse, *ibid*, at [107]; Lord Scott, *ibid*, at [161].
[166] *Cf Goldsworthy v Brickell*, above n 44 at 401 (Nourse LJ).

class was identical, and it was a device of substantive policy rather than true fact-finding. That the circumstances of an individual case might incidentally justify an inference of undue influence as having actually occurred *inter partes*, independent of artificial presumption—which, it might be noted, is a possibility in connection with status-based Class 2(A) relations of influence as well as fact-based Class 2(B) ones—ought not to obscure the overall fact that the presumption, while strictly superfluous in the instant case, might nevertheless continue to operate for its original, policy-based reasons. Indeed, in *Johnson v Buttress* Dixon J was alive to this point when he observed the occasional practice, even at that time, of some courts, when armed with 'the presence of circumstances which might be regarded as presumptive proof of express influence', and in connection with the Class 2(B) special relations of influence in particular, to treat the case as if it 'were not governed by the presumption but depended on an inference of fact'.[167] But for Dixon J, the existence of circumstances that might give rise to the suspicion that 'active circumvention' had been practised[168] was merely an incidental fact, and a 'cause why cases *which really illustrate the effect of a special relation of influence in raising a presumption of invalidity* are often taken to decide that express influence which is undue should be inferred from the circumstances'.[169]

C. On Burdens and Proof: From Persuasive 'Presumption' to Mere 'Permissible Inference'

If the House of Lords' conceptualisation of the relations or encounters that attract the principles of relational undue influence are equivocal as to their elemental fiduciary character, there is no doubt as to the evisceration of the conventional fiduciary rationale by the unambiguous demotion of the presumption, in the Class 2 cases, to a mere permissible (as opposed to a mandatory and controlling) inference. Drawing an analogy with the doctrine of *res ipsa loquitur* at common law, three Law Lords in *Etridge* considered the onus of proof to rest throughout on the complainant who was alleging undue influence, and that the effect of the 'presumption of undue influence' commonly said to arise in the Class 2 cases, involving so-called relationships of 'trust and confidence', was merely 'descriptive of

[167] Above n 15, at 135. In making these remarks Dixon J referred to Scrutton LJ's observation in *Lancashire Loans Ltd v Black* [1934] 1 KB 380 (CA) 404 that common law judges were inclined in this area 'to rely more on individual proof than on general presumption, while considering the nature of the relationship and the presence of independent advice as important, though not essential, matters to be considered on the question whether the transaction in question can be supported'.

[168] For example the presence of manifestly inadequate consideration in combination with the special antecedent relation between the parties.

[169] *Johnson v Buttress*, above n 15, at 136 (emphasis added).

a shift in the evidential onus on a question of fact'.[170] In other words, 'presumption' here means 'permissible inference' in the sense that the natural probative weight of the basic facts proven by the claimant as part of the burden of production ('relation of influence' and 'inexplicable transaction') is itself sufficient to discharge the claimant's persuasive burden in the absence of adequate counterproof by the defendant.[171] As soon as the defendant adduces countervailing evidence under his or her burden of production, however, it is a matter then of deciding whether the claimant has succeeded in meeting the ultimate persuasive burden 'on the balance of probabilities', that is, after the court has

> drawn appropriate inferences of fact upon a balanced consideration of the whole of the evidence at the end of a trial in which the burden of proof rested upon the plaintiff. The use, in the course of the trial, of the forensic tool of a shift in the evidential burden of proof should not be permitted to obscure the overall position.[172]

The court thus treats the defendant's burden as a mere 'evidential' onus or a 'provisional burden'[173]—a mere tactical requirement to adduce inferentially inconsistent evidence to affect the weight of the plaintiff's evidence—as opposed to a 'persuasive' or 'legal' burden that the defendant must discharge to the appropriate standard of proof so as to avoid summarily losing. The defendant is not, in other words, required to adduce evidence that would legally justify a finding against the presumed (inferred) fact.

The difference between what their Lordships in *Etridge* describe as occurring forensically in relational undue influence claims and how Dixon J in *Johnson v Buttress* (for example)[174] explains the operation of the

[170] Above n 6, [16]. See also *ibid*, at [107], [161].

[171] Lord Nicholls, *ibid*, at [14]. Hence the analogy to *res ipsa loquitur*, where the basic evidence is 'of such a nature as to afford in itself sufficient proof of negligence': *Fitzpatrick v Walter E Cooper Pty Ltd* (1935) 54 CLR 200 (HCA) 218 (Dixon J).

[172] *Etridge*, above n 6, [16].

[173] *Cf* AT Denning, 'Presumptions and Burdens' (1945) 61 *LQR* 379, 380.

[174] I do not want to create the impression that the court's view in *Johnson v Buttress* as to the evidential effect of the presumption is universally accepted. It is frequently unclear in the cases whether the court, in applying the presumption of undue influence, is implementing a rule concerning the burden of proof in the sense of the burden of persuasion or the burden merely of going forward with the evidence (the burden of production). One can readily locate decisions to either effect, but seldom is the distinction between the two views discussed in the case law. An exception (besides *Etridge*), which is contrary to the view that the legal burden shifts in relational undue influence claims, can be found in Sopinka J's judgment in *Geffen v Goodman Estate*, above n 16, at 242–3. Although Justice Sopinka accepted that the presumption of undue influence was one of law, he noted the division among courts and commentators as to the evidential effect of the presumption. His view, in obiter dicta, was that the matters that the defendant had to prove in rebutting the presumption of undue influence were merely factors that the court should consider in weighing the evidence and not an application of the legal burden of proof. In other words, no persuasive burden fell on the

presumption in such cases under the conventional fiduciary rationale is obvious. Under the traditional fiduciary account, the presumption of undue influence was a true legal presumption, with compelling effect,[175] and not a mere 'presumption of fact' by which a court was permitted (but not required) to deduce the 'presumed' fact, logically and reasonably, from the basic facts once they had been proven to the satisfaction of the court. The presumption did not cease to control as a makeweight in the assessment of the evidence unless the court positively believed the opposing evidence to the appropriate standard. The analogy to *res ipsa loquitur* in *Etridge* would, therefore, be inapt if relational undue influence were a fiduciary rule, for the principle lacks the mandatory effect that distinguishes true presumptions from mere permissible inferences.[176] Moreover, as mentioned at the outset of this article, the force of the presumption in the Class 2 cases was stronger than the basic facts needed to trigger it. Unlike the principle of *res ipsa loquitur* in common law negligence, the basic facts, if unchallenged, did not have to preponderate of undue influence on their own. Indeed, to the extent that *res ipsa loquitur* can be considered a 'presumption' at all (so ignoring, *arguendo*, its lack of mandatory effect), it is a very different sort of presumption, having a very different legal effect, than the one traditionally associated with 'presumed' undue influence. *Res ipsa loquitur* is a 'probability presumption',[177] in the sense that it arises from 'human experience or tendency' and is based on

defendant in an undue influence claim; rather, his or her burden was merely an evidential one, albeit an evidential burden that was subject to 'the familiar maxim that in applying the standard of proof all evidence is to be weighed in light of the gravity of the issue to be decided': *ibid*, at 243, citing *Blyth v Blyth* [1996] AC 643 (HL) 673 and *Hornal v Neuberger Products Ltd* [1957] 1 QB 247 (CA) 266). Thus, the evidence in undue influence cases 'is subjected to close scrutiny and the cases indicate that before coming to a conclusion, the court must act only on clear and convincing evidence': *ibid*, at 243. The debate is a difficult one that cannot be resolved within the confines of this article. Indeed, many experts on evidence law vigorously contend that it is inaccurate in any circumstance to speak of a shifting of the legal burden (on a single issue) in the course of a trial. Others, however, suggest that it is arguable that certain presumptions do have this effect. See JD Heydon (ed), *Cross on Evidence*, 7th edn (Sydney, Butterworths, 2004) [7225], [7300]–[7310].

[175] That is to say, success on the basic facts compelled the operation of the presumption as a matter of law.

[176] 'The principle expressed in the phrase *res ipsa loquitur* does no more than furnish a presumption of fact': *Fitzpatrick v Walter E Cooper Pty Ltd*, above n 171, at 219 (Dixon J). It is merely the application of inferential reasoning and does not, like true presumptions, have mandatory effect. *Cf Schellenberg v Tunnel Holdings Pty Ltd* (2000) 200 CLR 121 (HCA) [22] (Gleeson CJ and McHugh J). I have in the past described the House of Lords' analogy to *res ipsa loquitur* in this context as a 'fair but limited one': Bigwood, above n 17, at 439, fn 28; Bigwood, above n 40, at 385–86, fn 69. On further reflection, I no longer consider it to be fair at all.

[177] Here I draw on LJ Cohen, 'Presumptions According to Purpose: A Functional Approach' (1981) 45 *Albany Law Review* 1079, 1092–93, who argues that presumptions can be classified into three major types, according to the purpose for which each was created: 'procedural presumption', 'probability presumption', and 'policy presumption'. Needless to

the strong practical likelihood or high probability that if the basic facts are shown, the presumed fact also exists.

The presumption of undue influence, in contrast, has traditionally functioned as a 'policy presumption,' in the sense that it results not simply from human experience or tendency as a matter of realistic possibility (let alone probability),[178] but rather, and more significantly, from 'societal predilections toward the existence of the presumed fact'.[179] As a policy presumption reflecting societal values, therefore, the presumption of undue influence received, because it deserved, a higher priority than those presumptions that merely reflected common sense, experience or human tendencies (such as *res ipsa loquitur*). As the strongest possible form of presumption it thus functioned, again unlike *res ipsa loquitur*, to shift the burden of proof (or the risk of non-persuasion), and not merely the evidential burden or burden of production, onto the other party. With policy presumptions, the court must always find the presumed fact unless persuaded by a preponderance of the evidence, or, in the fiduciary context, high-probability ('clear and convincing,' 'cogent and compelling') proof,[180] that the presumed fact does not exist.[181] The presumption's purpose thus determined the treatment it received in terms of: (a) the cogency of the basic facts that triggered the presumption (basic facts that disclosed a realistic possibility of abuse as opposed to the substantial likelihood of it); (b) the presumption's legal effect in creating a persuasive burden upon the defendant to supply the want of necessary evidence; and (c) the standard of proof that the defendant then had to meet in supplying that necessary evidence ('clear and convincing', 'cogent and compelling', as opposed to merely 'preponderating' proof).

Returning to *Etridge*, the analogy of the presumption of undue influence to *res ipsa loquitur* is simply asserted by the House of Lords. Not only is its 'inexplicability' test too high in terms of the cogency of the evidence, or the level of risk, considered necessary to activate the presumption in the Class 2 cases (concerning, in particular, what counts as a 'suspicious transaction'

say, not many presumptions fall neatly within one class, for many are founded, at least in part, on all three bases: efficiency or expedience, quantifiable human experience or tendencies, and social predilections or values.

[178] As Cohen, *ibid*, at 1093 points out, policy presumptions 'differ from probability presumptions because the relationship between the basic and presumed facts cannot be quantified'.

[179] *Ibid*.

[180] The counterproof must compel a finding against the presumed fact, so that 'in-between' evidence will not suffice. Higher standards of proof are required when particularly important individual interests are at stake, as demonstrated, eg, in criminal (and indeed fiduciary) law. *Cf* C Manolakas, 'The Presumption of Undue Influence Resurrected: He Said/She Said is Back' (2006) 37 *McGeorge Law Review* 33, 40.

[181] In other words, the defendant must overcome the presumption, not merely meet or balance it with counterproof so as to then leave it as a matter for the tribunal of fact to decide.

in the context of a special relation of influence),[182] the Law Lords omit to consider the various purposes that underlie different presumptions, where one size clearly does not fit all.[183] Presumptions are created to accomplish many different objectives,[184] so in truth '[n]o general rule can … be laid down as to the effect of a particular presumption in the actual trial of a case, for this depends upon the purpose it is designed to serve'.[185] Certainly there is no direct reference in the various *Etridge* judgments to the generic policy foundations of the presumption of undue influence, and so no attempt is made by any of the Law Lords to demote the traditional presumption on rational grounds (for example by showing that the policy considerations that originally justified the equitable device are obsolete, harsh, impractical or unjust).

Some Law Lords clearly had in mind the Class 2(B) cases when making the analogy to *res ipsa loquitur*, rather than the Class 2(A) cases, which were strangely singled out for separate treatment by Lord Nicholls and Lord Scott in particular. I say 'strangely' here because, historically, from the standpoint of the presumption of undue influence, both types of case—Class 2(A) and 2(B)—functioned in the same way. The remarks of Lord Nicholls and Lord Scott in that regard, however, are confused (or at least confusing), which again casts doubt on the extent to which the prior law, which they were abandoning *sub silentio*, was appreciated, at least in light of the authorities as they stood in the United Kingdom before 1985. Consider, for example, the following passage from Lord Nicholls' judgment:

> The evidential presumption discussed above [in the 'relationship' cases] is to be distinguished sharply from a different form of presumption which arises in some cases. The law has adopted a sternly protective attitude towards certain types of relationship in which one party acquires influence over another who is vulnerable and dependent and where, moreover, substantial gifts by the influenced or

[182] To be sure, Lord Nicholls in *Etridge*, above n 6, [25], [29], [30] favoured Lord Scarman's strict suspicion of abuse test in *Morgan*, above n 30, at 709—namely, whether the impugned transaction was 'explicable *only* on the basis that undue influence has been exercised to procure it' (emphasis added)—which is a departure from the fiduciary account to the extent that it appears to implement 'probability' regulation rather than 'possibility' regulation. For the difference between the two forms of risk regulation, see n 58 above.

[183] Cohen, above n 177, at 1091: 'Because presumptions are formulated to achieve different purposes, it makes little sense to adhere to a rigid, one-rule treatment policy for all presumptions.' See also K Broun, 'The Unfulfillable Promise of One Rule for All Presumptions' (1984) 62 *North Carolina Law Review* 697.

[184] Edmund Morgan once described the 'most intellectually satisfying solution' in the field of presumptions thus: 'The effect of the establishment of the basic fact of a presumption should depend upon the weight of the reasons which caused its origin or justify its persistence. The reasons which call presumptions into being and the purposes they are designed to achieve are various': EM Morgan, 'Further Observations on Presumptions' (1943) 16 *Southern California Law Review* 245, 250–51.

[185] *O'Dea v Amodeo* 170 A 486, 487 (Conn Sup Ct Err 1934).

> vulnerable person are not normally to be expected. Examples of relationships within this special class are parent and child, guardian and ward, trustee and beneficiary, solicitor and client, and medical adviser and patient. In these cases the law presumes, irrebuttably, that one party had influence over the other. The complainant need not prove he actually reposed trust and confidence in the other party. It is sufficient for him to prove the existence of the type of relationship.[186]

There are difficulties with this passage, which to my mind cast doubt on the soundness of Lord Nicholls' understanding of the Class 2(A) cases and their (traditional) relationship to the Class 2(B) cases. Although it is recognised, correctly, that there are dual presumptions operating in the Class 2(A) cases—first a presumption *of influence* based on proof of a status-based relation (solicitor–client, guardian–ward, and the like), and second a presumption *of undue influence* triggered by proof of a transaction calling for explanation within the context of that relation—and that those two presumptions are to be 'distinguished sharply' from each other, Lord Nicholls seems to transpose the true marks of distinction. He suggests, on my reading of the above passage, that the second presumption is identical to the 'evidential presumption discussed above' (a permissible inference only), and that the first presumption somehow arises as a result of 'a sternly protective attitude towards certain types of relationship in which one party acquires influence over another who is vulnerable and dependent' (albeit in combination with an inexplicable transaction). In fact, as demonstrated by Dixon J's judgment in *Johnson v Buttress*, the law's 'sternly protective attitude' explains entirely the sequent presumption of undue influence in the relationship cases, both Class 2(A) and 2(B), and is not demonstrated at all in and by the initial presumption merely of influence in the status-based Class 2(A) cases.

In terms of the presumption types discussed above, the first presumption is at best a common-experience 'probability presumption' (a provisional presumption of fact),[187] used to satisfy the legal burden on the claimant to prove merely one of the basic facts, whereas the latter is a much stronger 'policy presumption' (a compelling presumption of law) serving a very different function and having an entirely different legal effect.[188] That Lord Nicholls suggests that the presumption of the first type is 'irrebuttable,' allowing the court no freedom to draw a different conclusion than the one required, regardless of the strength of the opposing evidence as to the truth

[186] *Etridge*, above n 6, at [18].

[187] *Clarke v Hawke* (1865) 11 Gr 527, 544.

[188] This is exemplified by the fact that as social views of relationship-types change, so will the presumption. In other words, status-based relations of influence are very likely to contract in proportion to the changing roles and capacities of the parties to those relations in fact. See, eg, Lord Evershed MR's remarks in relation to 'the position of women in modern society' in *Zamet v Hyman* [1961] 3 All ER 933 (CA) 938 (concerning engaged couples).

of the matter, defies both reason[189] and prior authority.[190] It also contradicts his express approval earlier in his judgment[191] of Sir Guenter Treitel's defensible remark[192] that the question in the 'relationship' cases is whether one 'party has reposed sufficient trust and confidence in the other, rather than whether the relationship between the parties belongs to a particular type'.[193] With respect, this is all symptomatic of a discomforting instability that infects much of the doctrinal elaboration in the various *Etridge* judgments, including that which eventually leads to an implicit, but I think inevitable, abandonment of the fiduciary approach to relational undue influence.

As mentioned above, Lord Scott thought that the Class 2(A) cases were useful in identifying particular relations where the presumption arose, but he doubted the utility of the Class 2(B) classification.[194] He viewed the Class 2(B) cases as examples of cases where, in proving the relationship between the dominant and subservient parties, in combination with whatever other evidence could be mustered at that time, the claimant had afforded sufficient primary evidence to justify a finding of undue influence in the absence of contrary evidence on the other side—hence the analogy to the *res ipsa loquitur* principle at common law.[195] But as was also mentioned, and as Dixon J in *Johnson v Buttress* explained, this does not mean that such cases are not otherwise capable of being based on the true,

[189] First, the idea of a presumption being 'irrebuttable' means that it is not really a 'presumption' at all; rather, it can only be a substantive principle expressed in the language of a presumption. Secondly, what reason could exist for inferring, conclusively, interpersonal influence when such influence patently did not exist in fact? *Cf* Finn, above n 47, at 85.

[190] Courts have, in fact, admitted evidence to controvert the presumption of influence in the Class 2(A) category of case. See, eg, *Westmelton (Vic) Pty Ltd v Archer* [1982] VR 305 (SC) where the presumption of influence was rebutted in a solicitor–client case. *Cf* also *Goldsworthy v Brickell*, above n 44, at 401 (Nourse LJ) holding that influence is presumed in the Class 2(A) relationships 'unless the contrary is proved'; *Geffen v Goodman Estate*, above n 16, at 221 (Wilson J): 'Equity has recognized that transactions between persons standing in certain relationships with one another will be presumed to be relationships of influence *until the contrary is shown*' (emphasis added). Lord Nicholls in *Etridge* seems simply to blindly adopt a point made *per incuriam* by Stuart-Smith LJ in the Court below, above n 148, at [6].

[191] *Etridge*, above n 6, [10].

[192] I have also argued that the Class 2(A) presumption of influence should be abandoned in favour of proof of an actual relation of influence in each case, while retaining the policy foundations of the presumption of undue influence that follows upon proof of a suspicious transaction occurring within the scope of the factual relation of influence: see Bigwood, above n 40, at 433–4. In the wider fiduciary context, Flannigan has argued that, in identifying fiduciary status, the formal relationship ought at best to perform a 'modest signaling function' only, and that it 'may be preferable to detach ourselves from our remaining dependence on the status ascription of fiduciary responsibility, and move to a fact-based limited access test for all cases': Flannigan, above n 48, at 228–9. See also R Flannigan, 'The [Fiduciary] Duty of Fidelity', above n 49, text following fn 76.

[193] GH Treitel, *The Law of Contract*, 10th edn (London, Sweet & Maxwell, 1999) 380–81.

[194] *Etridge*, above n 6, [161].

[195] *Ibid*.

policy-inspired presumption. Rare will be the case where a claimant relies exclusively on the presumption for his or her proof, submitting no other evidence. A party who claims to have been victimised by relational undue influence will naturally attempt to adduce as much evidence as possible to substantiate the claim, so seldom (if ever) will a court be asked to decide such a claim on a set of facts that presented nothing but the bare relation and the impugned transaction, and excluded all other evidentiary facts and circumstances of probative inferential worth. This is true of both the Class 2(A) and the Class 2(B) claims. Often the claimant will be able to carry his or her burden of production so well in proving a Class 2(B) relation of influence and contextually suspicious transaction that the evidence goes significantly beyond the basic evidentiary facts that ordinarily would have sufficed to trigger the policy-based presumption. Often such basic facts will themselves supply sufficient, but not necessarily conclusive, evidence of an abuse of special influence in fact (if not of some independent exculpatory reason such as duress or unconscionable dealing), so as to support an inference of undue influence as a matter of logic and reasoning, independent of invocation of the artificial presumption, thus allowing the claimant to make out a 'prima facie case' of undue influence and to meet his or her burden of persuasion, without the presumption, if there is too little inferentially inconsistent evidence produced on the other side. In such cases the same result can be reached without the aid of the presumption, making its operation strictly redundant, but these cases do not remove the justification for the presumption, or give reason to deny its operation in the particular case; they merely remove the practical need for it in the particular case (and perhaps in most, but certainly not all,[196] Class 2(B) cases).[197] The presumption continues to function, albeit merely theoretically and superfluously, in all such cases.[198] To my knowledge, the criteria for activation of the policy-based presumption have never been defined in a

[196] It is unlikely, eg, that *Johnson v Buttress* would have been decided in favour of the complainant under the *Etridge* evidential approach, unless one accepts Starke J's view of the facts in that case. See n 85 above and accompanying text.

[197] See CB Mueller and LC Kirkpatrick, *Evidence*, 2nd edn (New York, Aspen Law & Business, 1999) § 3.4: 'In effect, this kind of inference is the residue of a presumption, and [the tribunal of fact may] draw the conclusion that the presumption would otherwise require.'

[198] Perhaps, therefore, 'redundancy' might be a ground for supporting demotion of the presumption and elimination of the Class 2(B) category (as argued, eg, in a note, 'Undue Influence in Intervivos Transactions' (1941) 41 *Columbia Law Review* 707), but their Lordships in *Etridge* do not make this argument at all, at least directly. Anyway, unless one thinks that *Johnson v Buttress* was wrongly decided, that case tends to demonstrate that the Class 2(B) category is not always redundant, as *Johnson v Buttress* presumably could not have been decided the same way under the *Etridge* forensic approach. The majority of the court in *Johnson v Buttress* admitted that there was insufficient evidence to establish undue influence as a positive fact: see n 86 above. As mentioned in the text above, however, the views expressed in *Etridge* about the disutility of the Class 2(B) category are asserted rather than fully reasoned.

way that excludes the possibility of concurrent operation of a permissible inference or probability-based presumption if the primary evidence supports it.

IV. CONCLUSION

Contract lawyers are accustomed to courts killing off, or at least downplaying, 'presumptions' in their field.[199] As McHugh J observed in *Commonwealth v Amann Aviation Pty Ltd*, a reliance damages case:

> the history of the law of evidence has seen an increasing rejection of presumptions and other artificial forms of reasoning in favour of allowing tribunals of fact to give such probative force to evidentiary materials as they think fit having regard to all the circumstances of the case.[200]

It seems that the House of Lords in *Etridge*, for the United Kingdom at least, has now consigned relational undue influence to that fate. Although they do not explicitly reject the fiduciary character of (at least some of) the relations that attract the equitable jurisdiction, the Law Lords are unambiguous in their view of the so-called 'presumption' of undue influence, and how it operates in the Class 2 line of cases. That view is incompatible with relational undue influence being maintained as a category within the framework of fiduciary regulation. Disappointingly, that view was asserted rather than fully reasoned. The conventional fiduciary rationale was not openly considered and then rationally rejected. The traditional jurisprudence appears simply to have been ignored. Assertion somehow became the substitute for analysis. The rejection was *sub silentio*. There was certainly no frontal and convincing attack on the judicial intuition or substantive policy that originally justified strict fiduciary regulation in relational undue influence cases.

That said, relaxation or abandonment of relational undue influence as a fiduciary rule might be defensible on rational grounds—for example, because the rule can be shown to be obsolete, superfluous or impractical;[201] or

[199] Notably in the area of intention to create legal relations. See, eg, *Fleming v Beevers* [1994] 1 NZLR 385 (CA); *Ermogenous v Greek Orthodox Community of SA Inc* (2002) 187 ALR 92 (HCA) [26] (Gaudron J, McHugh, Hayne, and Callinan JJ).

[200] (1991) 174 CLR 64 (HCA) 166. This observation was in connection with the suggestion, accepted in the court below, that the law presumed that expenditure in performing a contract would be recouped through such performance, so that a court was entitled to infer that the plaintiff's loss for the purpose of the assessment of damages was no less than the amount of that expenditure.

[201] See, eg, the note at (1941) 41 *Columbia Law Review* 707, 722: 'the relationships relied upon to raise the presumption are essentially ambiguous, pointing at least as readily to a valid reason for the favoritism exhibited in the particular transaction as to the conclusion that something underhanded has taken place. It is therefore difficult to see any justification for the raising of a presumption of "undue influence"'.

'paternalistic';[202] or to operate unfairly, harshly or unjustly;[203] or to deprive trial judges of the valuable ability to assess and weigh the evidence so as to get the truth of the matter;[204] or to be an 'incubator in which fraud may flourish and grow';[205] or to create too-ready incentives for disgruntled or disappointed claimants to challenge objectively concluded transactions, often many years after the event or the death of the transferor.[206] Also, the *Etridge* view of undue influence is not inconsistent with the (sometimes) considered approaches to Class 2 undue influence of several courts in other common law legal systems.[207] Still, no such reasons are explicit in the demotion of the traditional 'presumption' of undue influence in *Etridge*. The law in that respect is presented there as if it had always been that way.

It is hoped that other Commonwealth appellate courts,[208] if and when they eventually confront the question of adopting (or not) the *Etridge* undue influence principles,[209] will respond reflectively, and not merely reflexively, to the issue. Such courts should not, in my view, relax or dismantle the strict fiduciary operation of traditional relational undue influence law without first credibly responding to the conventional rationale that existed for the strict accountability regime in first place, that is, without furnishing plausible reasons why we should now comfortably reject the original premises for the strict regulation. Without gainsaying the possibility of such reasons coming forward, I would myself be surprised if the risk of, and hence judicial concern for, the mischief of opportunism in limited-access arrangements were any less demonstrable today than in the past. Surely those who concede deferential trust in others, or who are otherwise subordinate within a limited-access relation of influence, are as

[202] CP Reed, 'Comment' (1984) 18 *Law Teacher* 132, 134.

[203] See, eg, BD Stapleton, 'The Presumption of Undue Influence' (1967) 17 *University of New Brunswick Law Journal* 46, 64–5; A Craig, 'Evidential Presumptions' (2002) 152 *New Law Journal* 217, 218. Although fiduciary rules may operate harshly in some cases, this is offset by the fact that it is very easy for a fiduciary to insulate himself or herself, *ex ante*, against that risk, that is, by securing the obligee's permissive and transactional consent. A failure to do so might itself be indicative, at least of transactional neglect. On the concept of 'transactional neglect,' see R Bigwood, 'Contracts by Unfair Advantage: From Exploitation to Transactional Neglect' (2005) 25 *OJLS* 65.

[204] *Re Estate of Carpenter* 253 So 2d 697, 703–704 (Fla 1975).

[205] Stapleton, above n 203, at 64.

[206] See, eg, the note at (1941) 41 *Columbia Law Review* 707, 722: 'It seems, too, that instead of being the shield of a trusting, weakminded grantor, as it was intended to be, the "presumption" can easily become the sword of a disgruntled, disappointed heir.'

[207] For example, *Boardman v Lorentzen* 145 NW 750 (Wis 1914); *Re Estate of Carpenter*, above n 204, a case involving the execution of a will, but the principles were later held to apply equally to *inter vivos* gift transactions; *Majorana v Constantine* 318 So 2d 185 (Fla Ct App 2d Dist 1975); *Re Estate of Wood*, 132 NW 2d 35 (Mich 1965); *Geffen v Goodman Estate*, above n 16, at 242–3 (Sopinka J in obiter dicta).

[208] Though perhaps Australia, Canada, and New Zealand especially.

[209] In contrast to the separate principles relating to third-party disability in the enforcement of suretyship transactions, which is the main subject of the *Etridge* appeals.

vulnerable to exploitation today as they ever were. On the other side, it is doubtful whether twenty-first-century human beings, when tempted by ready opportunities to make secret personal gains, are generally any less likely to succumb to the self-regarding impulse than their nineteenth-century counterparts. Indeed, in the wider fiduciary setting it has been argued that 'there does not appear to be a plausible basis for sending a new signal that the regulation of opportunism [in limited-access arrangements] should now be scaled back'.[210] Similarly, for undue influence, it might be argued that allowing the presumption of undue influence to affect only the burden of production and not the burden of persuasion is problematic on the ground that the presumption itself embodies strong policy preferences that are not adequately served if only the burden of production is affected.

The ultimate effect of *Morgan* and *Etridge*, undoubtedly, is to enlarge the scope of undue influence so that it now potentially applies to all acts of unfair persuasion in *inter vivos* transactions, regardless of the relational context within which that persuasion occurs. To my mind this has rendered undue influence hollow as an independent doctrinal category. Stripped of its fiduciary underpinnings, relational undue influence is now practically and intellectually indistinguishable from other exculpatory categories or pleas in avoidance that have, in other legal systems, tended to regulate unfair persuasion or victimisation in arm's-length transactional encounters. Those categories or pleas are certainly capacious enough to absorb undue influence's historical burden. In those other legal systems the expansion of the duress and unconscionable dealing doctrines, for example, has tended to eliminate the need for a broader application of the undue influence concept, so it is possible that relational undue influence may, despite *Etridge*, continue to retain its distinctive character in some English-based legal systems outside of the United Kingdom. The possible elimination of the Class 2(B) category of undue influence in particular, though, seems to leave little room for an understanding of relational undue influence apart from unconscionable dealing, the criteria of which seem perfectly well suited to the administration of many undue influence claims.[211] Yet to the extent that both doctrines can be understood as legal devices for the

[210] Flannigan, above n 54, at 212.

[211] As Mason J stated in a leading unconscionable dealing case (*Commercial Bank of Australia Ltd v Amadio*, above n 11, at 461): 'There is no reason for thinking that the two remedies [of unconscionable dealing and undue influence] are mutually exclusive in the sense that only one of them is available in a particular situation to the exclusion of the other. Relief on the ground of unconscionable [dealing] will be granted when unconscientious advantage is taken of an innocent party whose will is overborne so that it is not independent and voluntary, just as it will be granted when such advantage is taken of an innocent party who, though not deprived of an independent and voluntary will, is unable to make a worthwhile judgment as to what is in his best interest.'

protection of vulnerable parties, it cannot be said that victims of unconscionable dealing and victims of relational undue influence are 'vulnerable' (to victimisation) in quite the same way. 'Specially disadvantaged' parties under the unconscionable dealing jurisdiction are typically inept, weak or unable to advance their best interests when entering into voluntary or consensual transactions with others who are much stronger, but that is not because they have renounced playing for advantage themselves in such transactions. They would or might press for advantage or pursue self-interest *inter se* if only they could, but *ex hypothesi* their special relative disability inhibits them from doing so on the occasion in question. Such persons are disadvantaged, but they are not 'exposed'. Deferentially trusting parties, in contrast, characteristically have surrendered, partially or completely, control in their decision-making to another and so are susceptible to a much greater extent and to a higher order of wrongdoing altogether. They are, to the extent of their surrender, truly exposed, and their exposure is to no less than betrayal or treachery. The corollary of such greater vulnerability and risk on the one side must be greater obligation on the other, which is *fiduciary* obligation. We may question, therefore, whether such parties would be adequately protected—their interests properly served—if their petitions for exculpation from impugned transactions were consigned to administration through what is effectively an unconscionable dealing inquiry only. That, I believe, is the ultimate and lamentable effect of *Morgan* and *Etridge*.[212]

[212] Interestingly, in *Lawrence v Poorah* [2008] UKPC 21, Lord Walker, on behalf of the Judicial Committee, observed obiter dicta: 'It is sufficient to say that the doctrines of undue influence and unconscionable bargain share a common root—equity's concern to protect the vulnerable from economic harm—but they are generally regarded as distinct doctrines … In particular, although the doctrine of unconscionable bargain involves the exploitation of the plaintiff's vulnerability, it does not depend on a pre-existing relation of actual or presumed *confidence*. The doctrine of unconscionable bargain appears to be particularly vigorous in Australian jurisprudence …' (emphasis added).

Index

Introductory Note

References such as '178–9' indicate (not necessarily continuous) discussion of a topic across a range of pages. Wherever possible in the case of topics with many references, these have either been divided into sub-topics or only the most significant discussions of the topic are listed. Because the entire volume is about the law of contract, the use of this term (and certain others occurring throughout the work) as an entry point has been minimized. Information will be found under the corresponding detailed topics.