Penalties Reworked:

The Rule Against Penalties Restated,

Justified and Refined

Carmine Conte

Keble College

DPhil

Michaelmas Term 2014

Approximately 99,200 words

ABSTRACT

Penalties Reworked:

The Rule Against Penalties Restated, Justified and Refined

Carmine Conte, Keble College

DPhil, Michaelmas Term 2014

An 'agreed damages clause' is a contractual provision whereby the parties agree on the damages one party must pay if that party breaches a term. A valid agreed damages clause, known as a 'liquidated damages clause', is enforceable. An invalid agreed damages clause, known as a 'penalty clause', is not. The 'penalty rule' is the legal doctrine that discriminates between the two types of clauses.

This thesis adopts a rights-based, interpretive and formalist methodology to examine the penalty rule as it has evolved, and exists today, in Anglo-Australian commercial law. The investigation proceeds in three parts. First, this thesis *restates* the law. It derives a principled account of the rule from the leading cases. Second, the thesis *justifies* the law as restated. It reveals a theory that rationally explains the penalty rule, providing good or legitimate reasons why the rule is as it is. Finally, this thesis *refines* the restated law in the light of its theoretical justification. It takes interpretive steps to improve the penalty rule, casting the rule in the best possible light.

This thesis concludes that a penalty clause is best understood as requiring one party, upon a default, to incur a detriment that is manifestly excessive in light of the clause's legitimate function. The refined penalty rule invalidates a penalty clause by rendering it partially unenforceable. Also, this thesis illustrates that the rule encapsulates a principle of preventing consensual civil punishment, and that corrective justice norms justify that principle. As such, the penalty rule ultimately instantiates corrective justice. Further, this thesis demonstrates that a classificatory, correctivist theory scores extremely well on several criteria for evaluating private law theories. Such a rights-based theory is highly transparent, powerfully coherent, and practically beneficial. It also displays a tight fit with the current law.

In short, this thesis concludes that, viewed in the best possible light, the penalty rule is certain, elegant and (correctively) just.

TABLE OF CONTENTS

Part Three: Refinement

TABLE OF CASES

1 Australia

2 Canada

3 New Zealand

4 United Kingdom

TABLE OF STATUES

1 Australia

2 United Kingdom

1 INTRODUCTION

1 Overview

In life, rare items are interesting. The private law is no different. The rule against penalties is a good example. It is unique, being the only doctrine that constitutes a 'blatant interference with freedom of contract'.[1] Hence, the rule excites considerable jurisprudential interest. Recently, judges and jurists have begun asking difficult questions. For example, the Australian High Court has queried whether the classical tests for identifying a penalty clause, developed in the early 20th century, should be reformulated.[2] Leading academics have gone further still. Some have begun asking whether there is any basis at all for the rule, and suggesting that possibly it should be eliminated.[3] What is the form of the rule today? Should it properly remain unchanged, be reformulated, or be abandoned? In each case, why so?

This thesis systematically and theoretically evaluates the penalty rule. It has five aims.

The first aim is modest. It is to identify and understand the penalty rule's proper scope and content. This goal will involve two things. For one, it will require creating a utile framework to comprehend this area of the law. For another, it will involve rigorously analysing the legal materials to make sense of them, and then shaping those materials to provide the best understanding of the penalty rule.

The second aim is also conservative. The goal is to classify the doctrine, which is a preliminary step to justifying it. In doing so, this thesis will seek to set the terms for a proper jurisprudential debate concerning the rule's continued existence.

The third aim is much more difficult. It is to identify the best theoretical justification for the penalty rule. This thesis must look for a theory that explains the penalty rule in a way that takes seriously the law's demand that its decisions are justified rationally, and justified in terms of the legal materials. In short, the thesis will seek to prove that the normative foundation underlying the proposed classification is true.

---

1 *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [44].

2 *Ringrow Pty Ltd v BP Australia Pty Ltd* [2005] HCA 71, (2005) 224 CLR 656 [12].

3 Edwin Peel, 'Unjustified Penalties or an Unjustified Rule Against Penalties?' (2014) 130 LQR 365, 365; Sarah Worthington, *Equity* (2nd edn, OUP 2006) 227.

The fourth aim is to refine the penalty rule with its proper theoretical justification in mind. The goal is to cast the law in the best possible light. The refined rule must be capable of better explaining the seminal past decisions; of providing a sound guide for judicial decision-making; and of clear commercial application.

The final aim is self-reflectively to evaluate the refined rule. This thesis will attempt to assess whether the improved penalty rule paints the best picture of the law that can be designed. Three questions must be answered. Is the refined rule normatively appealing? Is it theoretically satisfactory? Is it able successfully to function in practice? A rule that casts the law in the best possible light will answer each question affirmatively.

This thesis reaches three main conclusions, as follows.

To start, contrary to long-held judicial and scholarly views, the penalty doctrine is not *prescriptive*, mandating civil parties appropriately to mimic a judicial compensatory remedy. Rather, it is *proscriptive*, prohibiting the parties from punishing each other. As such, the penalty rule encapsulates a principle of preventing consensual civil punishment.

Next, corrective justice norms justify the preventing consensual punishment principle, therefore the penalty rule instantiates corrective justice. Pursuant to this form of justice, punishment is foreign to the private law of wrongs, and hence it contains a general rule against punitive damages. By analogy, punishment is out of place in the private law of consent, and hence it contains a rule against penalties. In fact, preventing consensual punishment is the a fortiori case.

Lastly, this thesis concludes that a penalty clause is best understood as a clause requiring one party, upon a default, to incur a detriment that is manifestly excessive in light of the clause's legitimate function. Conceptualised correctly, the penalty rule prevents a party from enforcing a penalty clause by rendering it partially unenforceable. By this small interpretive step, the rule is rendered more principled, and the private law obtains more potential to be understood as a conceptually unified system.

In reaching these conclusions, this thesis will present an argument that adopts a particular strategy. It may be referred to as a 'case law-focused reflective equilibrium'.[4] This strategy proceeds in three stages. First, it is necessary to examine the relevant case law and

---

[4] See Stephen A Smith, *Contract Theory* (OUP 2004) 11; Allan Beever and Charles Rickett, 'Interpretive Legal Theory and the Academic Lawyer' (2005) 68 MLR 320, 325; Allan Beever, *Rediscovering the Law of Negligence* (Hart Publishing 2007) 27; Carmine Conte, 'From Only the 'Bottom-up'? Legitimate Forms of Judicial Reasoning in Private Law' (2015) 35 OJLS (forthcoming).

derive a principled understanding of the law from it. Second, one must produce a theory that best explains the law. Importantly, no bright line divides the legal understanding and the theory production stages: the exercise requires a continual back and forth between the law and the theory. Finally, it is necessary to re-examine the law with that understanding and theory in mind with a view to refining the law, such as to produce a greater equilibrium between the law and the theory.

This thesis is divided into three parts. Each part corresponds to a discrete stage in the process of a reflective equilibrium.

Part One seeks to *restate* the law. This part aims to derive a principled account of the penalty rule from the leading cases. Thus, it will examine and attempt to understand a great deal of the substantive law. This part will seek to provide the best interpretation of the penalty rule based on the present authorities.

Part Two aims to *justify* the law as restated. This part will search for the penalty rule's proper theoretical justification. So, it will analyse the penalty rule's possible theoretical justifications. This part will attempt to reveal the theory that best explains the rule, and prove that the theory is true.

Part Three endeavours to *refine* the law in the light of its theoretical basis. This part will seek to reconcile the restated law with its proper theoretical explanation. Hence, with the restated law and proper justification in mind, this part will seek to re-examine the law and refine the penalty rule. It will also venture to apply the refined rule systematically to some of the seminal cases. This part will aim to take interpretive steps to improve the penalty rule, casting the rule in the best possible light.

The three parts of this thesis contain 11 chapters. A short outline of each chapter follows.

Part One comprises five chapters. To begin with, Chapter 2 provides a short historical overview of the penalty rule. It examines the doctrine's history to the degree necessary to highlight the rule's significance and meaning, thereby enhancing how the modern rule is understood. Then, Chapter 3 canvasses the orthodox view of the penalty rule. This view is labelled 'orthodox' because a 'consensus of those familiar with the law'[5] would adhere to it. For example, most legal textbooks take this view, and generally most commentators and practitioners assume it to be accurate. Next, Chapter 4 examines the penalty rule's

[5] Stephen A Smith, *Contract Theory* (OUP 2004) 9.

jurisdictional test. Two questions are asked, and answered. When will a court review an agreed damages clause? Are the jurisdictional limits properly set? Following this, Chapter 5 analyses the penalty rule's tests of validity. It evaluates each of the tests potentially used to determine whether an agreed damages clause is valid, and asks whether any test of validity is properly the sole or primary test. The chapter also addresses three problematic issues that arise when Anglo-Australian courts apply these tests. Turning to the final chapter in this part, Chapter 6 examines the consequences of applying the tests of validity. It looks at various issues. What happens if an agreed damages clause is deemed valid? What if it is declared 'penal'? In the latter case, does the penalty rule face any intractable problems?

Part Two includes three chapters. Chapter 7 comprehensively assesses the penalty rule's putative theoretical justifications. The leading justifications considered are: reigning in substantive unconscionability; promoting economic efficiency; and preventing 'overcompensation'. The chapter illustrates that the leading theories do not satisfactorily explain the rule. Chapter 8 reveals the penalty rule's proper theoretical explanation. It presents a classificatory, correctivist theory of the rule. The chapter demonstrates that this two-level theory *does* satisfactorily explain the law. Chapter 9 evaluates the classificatory, correctivist theory from a variety of perspectives. It illustrates that on several criteria for evaluating private law theories, the propounded theory scores extremely well. In any event, it performs better than any of the leading theories of the rule.

Part Three contains one chapter. Thus, Chapter 10 reconciles the law as restated with its true theoretical justification. The chapter proposes a refined rule aligned with the doctrine's true classificatory and correctivist basis. It explains how the proposed rule operates at each of the relevant stages of inquiry. It also illustrates how a rule derived from a reflective equilibrium—an equilibrial rule—would generate the same outcome as that reached in the leading cases, but through more elegant, unified and powerful reasoning.

For completeness, Chapter 11 contains a short conclusion.

One final introductory matter remains. It concerns how the remainder of this chapter is structured. It comprises four sections. First, the chapter outlines the scope of the thesis. This section explains that the thesis is limited to examining the general legal controls operating on Anglo-Australian commercial contracts. Second, the chapter defines several important terms. It also posits a paradigm scenario to aid the reader to understand the analysis that follows. Third, the chapter describes the methodology adopted. In short, this thesis takes a 'rights-based' approach. This section explains the key features of that approach. Finally, the chapter explains the concept of private law 'classification'. This thesis conducts a classificatory analysis of the penalty rule, based upon the Birksian private law taxonomy. Therefore, this section outlines, and expands on, that foundational scheme.

2 Scope

The penalty rule has existed for centuries. The private law in most Commonwealth jurisdictions contains such a rule. It has also inspired legislative developments. Accordingly, considered holistically, the penalty rule is vast. Any fruitful analysis must be confined. As such, this thesis contains two limits.

On the one hand, it contrasts *English* and *Australian* jurisprudence only. There are two reasons for this.

First, it is legitimate to treat Anglo-Australian law as one body of common law, pointing out the relevant differences where they exist, because English and Australian law substantially share basic legal doctrine and values. English and Australian lawyers 'speak a common language and share many legal values', whilst the core areas of Australian private law have 'English underpinnings … conceptual, analytical and cultural'.[6]

Second, for several years, most commentators have viewed English and Australian law as diverging on important issues concerning the penalty rule—such as the consequences of invalidity—though perhaps the two jurisdictions have recently begun to speak with one voice. Conversely, for many years, English and Australian appellate courts appeared united on other central issues—such as the relevant jurisdictional test—but now one country has seemingly changed direction. Thus, contrasting the present law in these two jurisdictions provides the most useful method of painting the best picture of the penalty rule. It permits us to ask: where each jurisdiction adopts a different approach, can the approaches be reconciled? If not, which view fits most easily with the general picture of the law?

On the other hand, this thesis considers only *general law* controls on *commercial* contracts. It does not discuss statutes regulating penalty clauses in specific types of agreements, such as consumer[7] or hire purchase contracts. The reason for this is twofold. For one thing, legislators' power to make law differs from that of judges. Legislators are democratically

---

[6] James Douglas, 'England as a Source of Australian Law: For How Long?' (2012) 86 ALJ 333, 335 (footnote omitted), 349.

[7] eg Unfair Terms in Consumer Contracts Regulations 1999, SI 1999/2083 sch 2 [1(e)]; Competition and Consumer Act 2010 (Cth) sch 2 s 25(1)(c).

elected, and pass laws that seek to express the different interests of all segments of a pluralist society. Judges are not; and do not make law for the same purpose. Accordingly, one must use different yardsticks to assess statute law and judge-made law. For another thing, in the statutory context one factor straightforwardly provides an intervening basis. The narrow legislative controls relate to particular types of contracts involving a situational inequality in bargaining position, and the potential for one party to exploit its dominant position. The same cannot be said about commercial contracts as a category.

One final point regarding scope remains. This thesis is only concerned with the penalty rule. It does not examine any of the doctrines of relief against forfeiture, as properly understood. Space constraints preclude a full explanation of this issue. Nevertheless, the following comments seem pertinent.

Seemingly there exist three central forfeiture doctrines in the private law: relief against forfeiture of deposits; relief against forfeiture of part payments; and relief against forfeiture of proprietary or possessory rights. Undoubtedly, the penalty rule is related to these doctrines, for the same fact pattern might provide a claimant with a claim under the penalty rule and for relief against forfeiture. However, although the matter is open to argument, as best interpreted the penalty rule is separate and distinct from each of the forfeiture doctrines. For example, the elements a claimant must prove to succeed in a claim for relief against forfeiture of a proprietary right differ greatly from those a claimant must establish to obtain relief under the penalty rule.[8]

A potential misunderstanding must also be addressed. As Chapter 4 notes, when discussing the penalty rule jurisdiction, some Anglo-Australian judges and commentators have used the language of 'forfeiture'. This does *not* mean that the penalty rule subsumes one or more of the forfeiture doctrines. Properly understood, the situation is this. In the penalty rule context, the courts have used the term 'forfeiture' to refer to the extinguishment of a right. In some cases, a claimant must establish a 'forfeiture' in this sense to succeed. Similarly, some or all of the forfeiture doctrines adopt the 'forfeiture' concept in the sense of an extinguishment. However, in the forfeiture doctrine context, the 'forfeiture' concept constitutes merely one of several elements that a claimant must prove in order to obtain relief. It follows that the penalty rule and one or more of the forfeiture doctrines *overlap*, in that each uses the 'forfeiture' concept to mean an extinguishment of a right,[9] and that it may form an element of a claim

[8] cf Chapter 3 and *Shiloh Spinners Ltd v Harding* [1973] AC 691 (HL) 723-24.

[9] This is not uncommon. For instance, 'detriment' arguably forms one element of a claim for promissory estoppel *and* for breach of confidence.

under the relevant rule. It does not follow that the penalty rule *sweeps up* all or any of the forfeiture doctrines.

3 Terminology

Imprecise terminology pervades the penalties field. For example, many commentators use the term 'penalty' to refer to *both* valid and invalid agreed damages clauses.[10] That will not do. Terminological inexactitude leads to conceptual confusion. Thus, important terms must be defined in a more discriminating way. The relevant terms are: 'damages', 'agreed damages clause', 'liquidated damages clause' and 'penalty clause'. They are discussed in order.

A 'Damages'

The meaning of 'damages' is controversial. Traditionalists say that it is a common law remedy, synonymous with compensation for loss.[11] However, in this thesis, 'damages' means a financial remedy for breach of a duty owed to another. Typically, the term refers to a court order to the defendant to pay money to the claimant.

Damages may be classified according to their function. If so, an adjective indicating the remedy's function will precede the word 'damages'. For example, 'compensatory damages' refers to a financial remedy for a wrong that seeks to repair a claimant's loss.

B 'Agreed damages clause', 'liquidated damages clause' and 'penalty clause'

As traditionally understood, an 'agreed damages clause' is a provision whereby the parties agree on the damages one party must pay if that party breaches a term. The status of such clauses is indeterminate. A *valid* agreed damages clause, known as a 'liquidated damages clause', is enforceable according to its terms. An *invalid* agreed damages clause, known as a 'penalty clause', is not. In this thesis, the 'Penalty Rule' refers to the legal doctrine that discriminates between these two types of clauses.

An agreed damages clause displays two central features.

---

[10] eg Law Commission, *Penalty Clauses and Forfeiture of Monies Paid* (Law Com WP No 61, 1975) [9].

[11] *Kuddus v Chief Constable of Leicestershire Constabulary* [2001] UKHL 29, [2002] 2 AC 122 [95].

The first is that an agreed damages clause *supplants a court's jurisdiction* to determine the innocent party's general damages.[12] Effectively, the parties 'step into the shoes' of the court for the purpose of awarding the damages payable for one party's breach.

The second is that the agreed damages amount is necessarily *quantifiable* at the contract date.[13] The agreed damages amount must be quantifiable in the sense of being fixed or inelastic, or calculable by arithmetic process (such as by using a formula). Accordingly, agreed damages differ from general damages, as the latter are not quantified until a court assesses the relevant amount. Further, as Chapter 3 explains, the Penalty Rule tests of validity are applied as at the contract date. Accordingly, although agreed damages clauses may provide a formula for assessment, a court must be able to determine the amount the formula produces without reference to facts occurring *after* the contract date. Thus, the amount must be quantifiable 'in a way specified in or ascertainable from the contract itself …'.[14]

Importantly, the supplanting and quantification features are definitional: if a clause does not display these features, then it is not an agreed damages clause at all. Consequently, the Penalty Rule simply does not apply.[15]

As stated, if a liquidated damages clause exists, the innocent party may enforce the provision according to its terms. As a result, each party will know in advance what the innocent party's right, and the guilty party's duty, will be upon breach. This provides certainty.[16] Further, if the clause is a liquidated *compensatory* damages clause, the innocent party may sue on the clause and recover the sum specified without having to prove that she suffered any loss.[17] Given this, unless the guilty party denies that he breached the contract, the clause will avoid risky, time-consuming and costly litigation.[18] Accordingly, the main advantages of liquidated damages clauses are that they provide certainty, save expense and avoid litigation. These are significant practical advantages, especially in the commercial sphere.

---

[12] *Murray v Leisureplay plc* [2005] EWCA Civ 963, [2005] IRLR 946 [52].

[13] Edwin Peel, *The Law of Contract* (13th edn, Sweet and Maxwell 2011) [20-130]; Andrew Burrows, *Remedies for Torts and Breach of Contract* (3rd edn, OUP 2004) 444; Keith Mason, JW Carter and Greg Tolhurst, *Mason and Carter's Restitution Law in Australia* (2nd edn, LexisNexis Butterworths 2008) [1824].

[14] *Rothenberger Australia Pty Ltd v Poulsen* [2003] NSWSC 788, (2003) 58 NSWLR 288 [27].

[15] *Lombard North Central plc v Butterworth* [1987] QB 527 (CA) 536.

[16] *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 193.

[17] *Boucaut Bay Co Ltd v Commonwealth* (1927) 40 CLR 98 (HCA) 106; *Robophone Facilities Ltd v Blank* [1966] 1 WLR 1428 (CA) 1447, 1449.

[18] *Robophone Facilities Ltd v Blank* [1966] 1 WLR 1428 (CA) 1447; *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 193.

D 'Paradigm scenario'

To assist with the exposition, this thesis throughout considers a 'paradigm scenario'. This situation provides 'the paradigm case'[19] in which the Penalty Rule is engaged, or the 'standard application'[20] of the rule. The situation is as follows.

Two parties, A and B, agree a commercial contract. That contract contains an agreed damages clause requiring B to pay a fixed sum of money to A upon B breaching a particular term, or set of terms.

4 Methodology

This thesis adopts a 'rights-based' methodology. This analytical approach seeks to understand and explain private law doctrine by focussing exclusively on the rights that the parties hold against each other, rather than on external influences potentially affecting the private law, such as community welfare goals. The fundamental claim of rights analysis is that Anglo-Australian private law 'is best viewed as concerned with the rights which, one against another, people are able to realize in courts'.[21]

As with any legal methodology, particular rights theorists may differ in how precisely they apply the rights-based approach. That said, generally speaking, the rights-based methodology contains four related, central characteristics: it is interpretive, formalist, non-instrumental and monist.[22] Each characteristic is discussed, in order, below.

A Interpretive

Smith suggests that four types of 'account' may be presented of any legal area: historical, prescriptive, descriptive and interpretive.[23] An historical account seeks to explain how and why the law has developed precisely as it has. A prescriptive account outlines what the law ought to be. A descriptive account seeks to depict the law at a particular period in time. Conversely, an interpretive account aims to enhance legal understanding by revealing an intelligible order in the law and highlighting its meaning. It seeks to reveal an intelligible order in the law by explaining why particular characteristics of the law are important, or not; and by

---

19 *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [46].

20 *Ringrow Pty Ltd v BP Australia Pty Ltd* [2005] HCA 71, (2005) 224 CLR 656 [10].

21 Andrew Burrows (ed), *English Private Law* (3rd edn, OUP 2013) ix.

22 See Donal Nolan and Andrew Robertson, 'Rights and Private Law' in Donal Nolan and Andrew Robertson (eds), *Rights and Private Law* (Hart Publishing 2012).

23 Stephen A Smith, *Contract Theory* (OUP 2004) 4-5.

identifying connections between those essential features.[24] In short, interpretive legal theory aims to 'provide the most plausible, coherent and appealing account of the law *as it stands*'.[25] Thus, an interpretive account draws on historical, prescriptive and descriptive accounts of the law—it displays elements of each of them—but it is not identical to any of them. In particular, though interpretivism may display normative features, it is not prescriptive in nature, for it seeks primarily to outline how best to understand the law as it *presently* exists; is constrained by the history and shape of the case law; and must adhere to 'law's self-understanding', that is, it must employ language, arguments, concepts and techniques of the same *kind* as those that courts typically adopt.[26]

The interpretive method seeks to provide the 'best' explanation of legal data. Explanatory force is the overriding concern. Accordingly, rights theorists are clearly interpretivists, as they seek to give the 'best' account of the private law, or of specific private law doctrines. For example, Stevens argues that rights-based analysis seeks 'to explain the law so that it makes coherent sense and to account for it in the best possible light'.[27] The goal, then, is to paint the best picture possible.

According to Smith, the criteria for assessing how much explanatory force a theory provides are fit, transparency, coherence and morality.[28] These criteria may be used to evaluate private law theories generally, including an interpretive theory. Many legal positivists would discard the morality criterion, as they would view as irrelevant a theory's moral appeal. This thesis takes no stand on whether that category is appropriate. Rather, this thesis focuses on the remaining three axes—fit, transparency and coherence—which it now considers in turn.

I Fit

A theory 'fits' if it 'is consistent with, and supported by, the … precedent based legal rules and … decisions'.[29] Accordingly, a theory fits if it generates outcomes that are consistent with the decisions made in the cases. However, two points must be noted. The first is that fit is not determinative, as precedents bind judges, not lawyer's understandings. The second is that

---

[24] Stephen A Smith, *Contract Theory* (OUP 2004) 5.

[25] Donal Nolan and Andrew Robertson, 'Rights and Private Law' in Donal Nolan and Andrew Robertson (eds), *Rights and Private Law* (Hart Publishing 2012) 6 (original emphasis).

[26] Stephen A Smith, 'Review: Taking Law Seriously' (2000) 50 UTLJ 241, 241, 255.

[27] Robert Stevens, 'Damages and the Right to Performance: a Golden Victory or Not?' in Jason W Neyers, Richard Bronaugh and Stephen GA Pitel (eds), *Exploring Contract Law* (Hart Publishing 2009) 198.

[28] Stephen A Smith, *Contract Theory* (OUP 2004) 7-32.

[29] Stephen A Smith, *Contract Theory* (OUP 2004) 7 (footnote omitted).

the importance of fit is relative to the nature of the case. Beever distinguishes four general categories of cases.[30] 'Central' cases are ones of utmost importance; they are foundational. 'Standard' cases are ones that broadly exemplify the principles that the central cases enunciate. 'Controversial' cases are consistent with the general body of case law, but seem to jar with the general law in some way. Lastly, 'divisive' cases are ones that are inconsistent with the general body of case law. The importance of a tight fit varies depending on the category of case, as follows: 'It is more important that the theory fit the central cases than the standard cases, the standard cases than the controversial cases, and the controversial cases than the divisive cases.'[31]

II Transparency

A theory is 'transparent' if it is able to 'take account … of the reasons that judges give for their decisions'.[32] In other words, a theory is transparent if it conforms to the courts' general reasoning. One can distinguish two levels of transparency.

A theory is *strongly* transparent if the explanation that it gives tightly matches the judges' reasoning in cases in the particular private law area being examined. For example, a strongly transparent theory of duress would explain that doctrine in terms of deficient consent and impaired autonomy. Although the theoretical explanation of a doctrine may be more abstract than the judicial one, the former must lead to the latter by logical steps: the theory must work 'through, rather than around, judicial reasoning'.[33] In short, the judicial reasoning must merely *instantiate* the theoretical explanation.

Conversely, a theory is *weakly* transparent if it respects the law's self-understanding, and adopts legal reasons of the same kind as courts typically adopt, but only matches the judges' reasoning in the instant area either moderately, loosely or not at all. For instance, a weakly transparent theory of mitigation would explain that doctrine in terms of 'good faith'. Again, the account may be more abstract than the standard judicial account, though it must explain the doctrine using abstract concepts that logically lead to concrete principles that a judge might adopt. In contrast to strong transparency, here the recognisably legal principles are ones that a judge *might* apply, even though only a few judges have, or no judge has, applied them to date.

---

[30] Allan Beever, *Rediscovering the Law of Negligence* (Hart Publishing 2007) 30.

[31] Allan Beever, *Rediscovering the Law of Negligence* (Hart Publishing 2007) 31.

[32] Stephen A Smith, *Contract Theory* (OUP 2004) 25 (footnote omitted).

[33] Stephen A Smith, *Contract Theory* (OUP 2004) 27.

III Coherence

Two notions of 'coherence' exist.[34]

Under the *weak* notion, a theory satisfies coherence to the extent that it shows the private law, or parts of it, to avoid self-contradiction and to be internally consistent. Appellate court judges see this as part of their mandate, frequently supporting this objective. For example, Heydon J has confirmed extra-judicially that judge-made laws should be 'substantially consistent' and 'harmonious, without anomaly or internal conflict. They should hang together in a connected fashion.'[35]

Pursuant to the *strong* notion, a theory is coherent if it shows that the private law, or parts of it, can be understood as a unified system, perhaps under a single, central theme. Again, this is orthodox judicial method. For instance, in *Customs and Excise Comrs v Barclays Bank plc*, Lord Rodger acknowledged that appellate courts function partly to '[boil] down a mass of case law' and distil 'some shorter statement of the applicable law' or identify 'some compact underlying rule …'.[36] Similarly, writing extra-judicially, McHugh J noted that it was a judicial duty to rationalise general principle 'with a view to bringing more unity and symmetry to the general law'.[37]

B Formalist

There are many types of legal formalism. In this thesis, 'legal formalism' refers to the view that it is the 'structural' private law concepts that determine the rule to be applied in, and the result of, particular kinds of cases. Formalism is an interpretive approach to the law that attempts to find and elucidate an intelligible order in the law by analysing its form. Put another way, a legal formalist 'seeks to gain theoretical information [and] practical guidance concerning law through attention to its "form" as opposed to its (historically and geographically variable) content …'.[38]

---

[34] Allan Beever, *Rediscovering the Law of Negligence* (Hart Publishing 2007) 22.

[35] JD Heydon, 'Limits to the Powers of Ultimate Appellate Courts' (2006) 122 LQR 399, 414. See also *Dietrich v R* (1992) 177 CLR 292 (HCA) 320.

[36] *Customs and Excise Comrs v Barclays Bank plc* [2006] UKHL 28, [2007] 1 AC 181 [51]. See similarly *Deutsche Morgan Grenfell Group plc v IRC* [2006] UKHL 49, [2007] 1 AC 558 [156].

[37] Michael McHugh, 'The Judicial Method' (1999) 73 ALJ 37, 40-41 (footnote omitted).

[38] Martin Stone, 'Formalism' in Jules Coleman and Scott Shapiro (eds), *The Oxford Handbook of Jurisprudence and Philosophy of Law* (OUP 2002) 170 (footnote omitted).

Rights theory is formalist in that its proponents have a distinct interest in the frameworks that underlie private law. For example, Beever argues that private law doctrines have a particular structure, and the 'first … task of legal analysis' is to reveal what that structure is.[39] Rights-based analysis emphasises several structural features of the private law. An important one is correlativity. We now consider three examples.

First, rights and duties correlate. So, Hohfeld famously argued that lawyers' use of the term 'right' was ambiguous, and that it in fact concealed four distinct types of legal entitlement: a claim, a privilege, a power and an immunity.[40] Where a party holds one of these four types of 'rights', a counterparty occupies a correlative, specific normative position: a duty, 'no-right', a liability and a disability respectively. The most common meaning of the word 'right' is the first one. This may be labelled a 'claim right', which correlates with a legal duty. Thus: 'X has a claim right that Y should do, or refrain from doing, an act if and only if Y has a duty to X to do, or refrain from doing, that act.'[41] A *primary* claim right correlates with a *primary* duty. For instance, X may hold a primary right against Y that Y performs a contractual obligation that she owes X. Where a defendant violates a claimant's primary claim right, a secondary claim right arises. A *secondary* claim right correlates with a *secondary* duty to repair. For example, X may hold a secondary right against Y for general damages for Y's breach of her contractual duty she owed to X.

Second, rights analysis links who can sue and who can be sued. It necessarily requires that that the only person who can sue is the right-holder, and the only person who can be sued is the person who violated the right.[42]

Finally, for rights theorists, it is one party's infringement of another party's primary right that defines the parties' legal relationship. The bilateral structure of this relationship links only a *specific* claimant with a *specific* defendant. This explains one of the private law's most basic features, namely, that a particular claimant may only enforce her claim right against a particular defendant who has breached his or her duty to that claimant.

C Non-instrumentalist

---

[39] Allan Beever, *Rediscovering the Law of Negligence* (Hart Publishing 2007) 515.

[40] WN Hohfeld, 'Some Fundamental Legal Conceptions as Applied in Judicial Reasoning' (1913) 23 Yale LJ 16, 30, 32.

[41] Robert Stevens, *Torts and Rights* (OUP 2007) 4.

[42] Robert Stevens, *Torts and Rights* (OUP 2007) 173, 244.

The rights method is characteristically non-instrumentalist, as it construes the private law as being internally intelligible and thus not needing to refer to external purposes. For example, Smith portrays rights-based contract theory as 'concerned with duties that contracting parties owe to each other rather than any broader social goal …'.[43] Accordingly, pursuant to rights theory, 'policy-based' reasoning is illegitimate in private law judicial decision-making. 'Policy' may be broadly defined as reasons supporting or rejecting a legal rule or outcome that concern community interests not related to the law's form. Thus, the category includes issues arising from social morality, economic efficiency and public opinion.

Rights theorists argue that policy considerations can have no role to play in judicial reasoning because these concerns are non-justiciable.[44] There are three reasons why policy is non-justiciable.

To start, rights-based scholars argue that judges lack the technical competence to assess the merits of policy arguments. So, in *Richardson v Mellish*, Best CJ held that courts 'have not the means of bringing before them all those considerations which ought to enter into the judgment of those who decide on questions of policy …'.[45] But even if courts could do so, they have not the expertise to evaluate those concerns. In particular, judges are not equipped to determine whether a policy-based decision will benefit the community, or not. Judges are not economists, philosophers, social scientists or statisticians.

Next, judges lack the political jurisdiction (and, hence, legitimacy) to determine policy claims. As Beever explains, when judges appeal to policy, they transgress their legitimate function, for they become 'unelected, unaccountable officials enforcing their personal political preferences on a sometimes unwilling public, apparently inconsistently with the notion of democracy'.[46]

Lastly, policy-based reasoning requires judges to weigh incommensurable considerations. Thus, conflicting policies often involve incommensurable concerns. Take economic efficiency and social morality, for instance. Suppose two parties enter a contract affected by illegality. Given the moral turpitude involved, it is more moral to invalidate the contract. But it is more efficient to enforce it. How is a court to determine, *in policy terms*, which outcome is better? It cannot. Stevens sums up the problem, 'If we ask which is the

---

[43] Stephen A Smith, *Contract Theory* (OUP 2004) 140-41.

[44] *McLoughlin v O'Brian* [1983] 1 AC 410 (HL) 430; *Hollis v Vabu Pty Ltd* [2001] HCA 44, (2001) 207 CLR 21 [117].

[45] (1824) 2 Bing 229, 242; 130 ER 294, 299.

[46] Allan Beever, *Rediscovering the Law of Negligence* (Hart Publishing 2007) 14.

greater of three kilos or six metres we are asking a question which it is impossible to answer as we have no common yardstick for [comparative purposes]'.[47] As these two policies are incommensurable, it is impossible to give reasons for choosing one over the other. Consequently, any decision that seeks to balance both concerns is necessarily unprincipled.

D Monist

Two views exist regarding the relationship between a right and a remedy. The first view is the 'monist' view. It considers that the right and the remedy are perfectly congruent. The monistic view treats the remedy as the mirror image of the right. As Birks famously said, the remedy is 'the same thing as the right, looked at from the other end'.[48] The second view is the 'dualist' view. It severs the remedy from the violated right. The dualist view believes that, in granting relief, a court 'makes a context-specific evaluation of the remedy that is most appropriate in all the circumstances'.[49] Accordingly, dualists support a form of 'discretionary remedialism'.

Rights theory is monistic in nature. Proponents view a remedy as justifiable only if it normatively 'matches' the right. Consequently, the nature and scope of the right dictates the remedial response to its infringement. There is barely any choice to be made. For example, Weinrib argues that rights theory 'treats the remedy as merely the continuation of the [claimant's] right in circumstances in which the defendant has, or has done, something inconsistent with that right'.[50] So, the right is the chrysalis, and the remedy is the butterfly that later emerges. Similarly, Stevens argues that a secondary claim right to general damages 'monetarises'[51] the primary claim right. In other words, the remedy is merely a financial quantification of the right that the claimant possessed at the start of the story.

From a rights perspective, then, 'discretionary remedialism' is unjustifiable. It fails to take seriously the concept of rights.

5 Classification

[47] Robert Stevens, *Torts and Rights* (OUP 2007) 310.

[48] Peter Birks, 'Definition and Division: A Meditation on Institutes 3.13' in Peter Birks (ed), *The Classification of Obligations* (Clarendon Press 1997) 24.

[49] Michael Tilbury, 'Remedies and the Classification of Obligations' in Andrew Robertson (ed), *The Law of Obligations: Connections and Boundaries* (UCL 2004) 17.

[50] Ernest J Weinrib, *Corrective Justice* (OUP 2012) 107-108.

[51] Robert Stevens, 'Damages and the Right to Performance: a Golden Victory or Not?' in Jason W Neyers, Richard Bronaugh and Stephen GA Pitel (eds), *Exploring Contract Law* (Hart Publishing 2009) 187.

This thesis undertakes a classificatory investigation of the Penalty Rule. So what is 'classification'?

Classification concerns mapmaking. It involves organising a mass of apparently unrelated legal claims and decisions into digestible, reasoned categories to allow lawyers usefully to compare the subject matter within each category, and across discrete categories, and to assist in judicial decision-making. In short, classification is about working out what are the various parts of the private law, and then making clear how they fit together. Every legal system must adopt *some* form of classification, for mapping is essential to organising legal thought.

Rigorous classification is essential for at least five reasons.

First, classification is important because how one labels and structures the private law shapes how one understands it. Just as a social scientist's understanding of the natural world is acquired primarily through classificatory schemes, so too a lawyer's understanding of the law is developed through categorical schemes. The only goal, then, is to aid understanding.

Second, it also ensures lawyers ask intelligible questions. As Lord Steyn has noted, 'classification is important. Asking the right questions in the right order reduces the risk of wrong decisions.'[52] Further, a categorical scheme provides a framework that can assist in clarifying the terms of a debate regarding unresolved analytical questions. Talking at cross-purposes aids no one.

Third, good classification increases consistency and coherence by enabling a judge to rationalise a group of cases and eliminate anomalies and contradictions.

Fourth, by providing the means to identify underlying reasons for similar treatment, it ensures that like cases are treated similarly, satisfying the 'uniformity principle'. The rule of law demands this.

Finally, classification is practically important. For example, categorisation prevents practitioners drowning in the 'information overload' of the modern age. The structure of the law is 'the software which allows the brain to keep the mass of information under intellectual control'.[53] If it is neglected, the price to be paid is a high one: private lawyers will not serve their clients well; and clients will lose faith in, and respect for, the law.

So, in order properly to understand and justify the Penalty Rule, it is necessary to

[52] *A-G v Blake* [2001] 1 AC 268 (HL) 290.

[53] Andrew Burrows (ed), *English Private Law* (3rd edn, OUP 2013) vii.

comprehend its place on the private law map. However, one can divide up the private law in a number of ways. Many maps are possible. In examining the rule, this thesis adopts and builds on the Birkisan private law taxonomy. This section outlines its features.

A The map

The premise underlying Birks's taxonomy is that a coherent framework of rational principle orders the private law. Accordingly, the writer's classificatory scheme seeks to find logical, rational order in the mass of decisions that a private law system produces. It does so because only transparent rationality can 'preserve the legitimised authority of the law … in a plural society'.[54]

For Birks, event and response formed a logical opposition. Thus, the Birksian taxonomy of private law obligations separates legal *causes* from legal *responses*. It seeks to divide the private law according to the particular event that generates a specific kind of right directly realisable in the courts. Importantly, the Birksian map brings order to the conceptual minefield of common law and equity by integrating the two into a list of events categorised according to rational criteria.

Essentially, the Birksian private law map posits that four causative events (consent, wrongs, unjust enrichment and 'other events') potentially generate one of five responses (compensation, restitution, perfection, punishment and 'other goals').[55]

The Birksian taxonomy is incomplete. Two non-miscellaneous responses are missing.

On the one hand, the Birksian grid does not list 'substitution' as a response. This is partly because Birks died before academic thinking regarding this response fully matured. Any events-based classification is not rationally defensible without it.

On the other hand, the table does not list 'disgorgement' as a response. Birks used the single label of 'restitution' to encompass both 'giving back' and 'giving up'. Although many say that the latter properly describes 'disgorgement', the writer saw no need for two labels. This is contestable.[56] English and Australian courts have held that a disgorgement remedy is neither compensatory nor restitutionary, but rather strips profits. Accordingly, judges view restitution

---

[54] Peter Birks, 'Rights, Wrongs, and Remedies' (2000) 20 OJLS 1, 22.

[55] Peter Birks, 'The Law of Restitution at the End of an Epoch' (1999) 28 UWAL Rev 13, 17; Peter Birks, 'The Law of Unjust Enrichment: A Millennial Resolution' (1999) Sing JLS 318, 322.

[56] See *Sempra Metals Ltd v IRC* [2007] UKHL 34, [2008] 1 AC 561 [180].

and disgorgement as discrete remedial functions.[57] In fact, a profit stripping remedy is only 'restitutionary' in the broad sense of being gain-based. But, if so, it is preferable to use the general term 'gain-based' to refer to the genus, leaving 'restitution' to encompass only a species. Imprecise terminology leads to confusion.

Accordingly, the Birksian private law taxonomy must be extended to include substitution and disgorgement as potential responses. The extended classification may be represented in a table, as follows.

Table A

Causative events

Responses	Consent	Wrongs	Unjust enrichment	Other events
Substitution				
Compensation				
Restitution				
Disgorgement				
Punishment				
Perfection				
Other functions				

Each header in the 'Causative events' columns denotes an events-based category, being a class of real world facts. These categories respond to the question: what events generate rights? Each header in the 'Responses' rows denotes a response-based category, being the function of a specific kind of right which is realisable in court. These categories respond to the question: what is the function of rights when they arise? So, the intersecting boxes reflect the function of a right that potentially may arise in reaction to a particular real world event. The court then replicates the relevant right in the court order. We now consider more closely each type of causative event, and each kind of response.

---

[57] *Harris v Digital Pulse Pty Ltd* [2003] NSWCA 10, (2003) 56 NSWLR 298 [414]; *Murad v Al-Saraj* [2005] EWCA Civ 959, [2005] WTLR 1573 [108].

B The causative events

A causative event is a recognised pattern of facts that happen in the real world. Four types of causative events generate rights implementing a particular response: consent, wrongs, unjust enrichment, and 'other events'. Each is considered in order.

Some rights arise from a manifestation of consent, meaning the expression of free will. A single person, or several persons acting together, may exercise consent. An example of the former is an express declaration of trust. An example of the latter is a contractual agreement.

Other rights arise from wrongs. A legal wrong is the breach of a duty owed to another.[58] However, 'a duty is merely the obverse of a right';[59] the right and the duty are 'the same thing, just viewed through different ends of the telescope'.[60] Accordingly, X's breach of his *duty* owed to Y is synonymously X's infringement of Y's *right* held against X. Further, a wrong is the infringement of a primary right that generates a secondary right.[61] Examples of wrongs include breaches of contract, equitable wrongs and torts.

Further rights arise from unjust enrichment. An unjust enrichment occurs where C without fault makes a reversible transfer of value to D, whereby D is enriched at C's expense in defined circumstances of injustice. For instance, an unjust enrichment will arise where C makes a spontaneous mistaken payment to D.

Finally, a variety of rights arise from 'other events'. The fourth category refers to miscellaneous events falling outside the other categories. It is a dustbin. Examples include the doctrine of necessitous intervention in another's affairs, tax obligations and obligations arising from court orders.

C The responses

A response refers to the function of a right that potentially arises in reaction to a causative event. The court replicates this pre-existing right by 'notavting' it: that is, the court extinguishes the original right and replaces it with a right arising from a court order.

---

[58] Robert Stevens, *Torts and Rights* (OUP 2007) 2.

[59] *Letang v Cooper* [1965] 1 QB 232 (CA) 247.

[60] NJ McBride, 'Rights and the Basis of Tort Law' in Donal Nolan and Andrew Robertson (eds), *Rights and Private Law* (Hart Publishing 2012) 342.

[61] *Photo Productions Ltd v Securicor Transport Ltd* [1980] AC 827 (HL) 848.

Under the Birksian scheme, responses are classified according to the function of the right. Pursuant to the extended classificatory scheme, these are: substitution, compensation, restitution, disgorgement, perfection, punishment and 'other responses'. The following examines the non-miscellaneous categories in sequence.

I Substitution

In recent years rights scholars have identified and developed a response called 'substitution'. Various forms of substitutive analysis exist. This thesis adopts Stevens's model.[62] Substitution seeks to vindicate an infringed right by substituting for the value of that right. Accordingly, where substitution responds to a wrong, the remedy is labelled 'substitutive damages'.

Stevens distinguishes *injuria*, being the wrong of the rights violation, from *damnum*, being the consequential loss arising from that infringement. Accordingly, the writer differentiates between substitutive damages and 'consequential damages as compensation for loss to the claimant, or gain to the defendant, consequent upon [the right's] infringement'.[63]

Under this approach, the claimant's basic award is for damages for the value of the infringed right. However, if the claimant suffers a loss as a result of the wrong, she may also claim consequential compensatory damages. Similarly, if the defendant makes a relevant gain by committing the wrong, the claimant may additionally claim consequential gain-based damages. So, in the breach of contract context, courts award substitutive damages for the value of the infringed right to performance (that is, the 'performance right'), and compensatory damages for any consequential financial or non-financial losses suffered.

Substitutive damages have four particular features.[64] First, they are assessed objectively. Second, the assessment date is the date of the infringement. What happens later is irrelevant.[65] Third, substitutive damages are awarded even if the claimant suffers no loss, or the defendant makes no gain, as a result of the rights violation. Finally, substitutive damages

---

[62] Robert Stevens, *Torts and Rights* (OUP 2007) 60-61; Robert Stevens, 'Damages and the Right to Performance: a Golden Victory or Not?' in Jason W Neyers, Richard Bronaugh and Stephen GA Pitel (eds), *Exploring Contract Law* (Hart Publishing 2009).

[63] Robert Stevens, *Torts and Rights* (OUP 2007) 60.

[64] Robert Stevens, *Torts and Rights* (OUP 2007) 60.

[65] eg *Joyner v Weeks* [1891] 2 QB 31 (CA) 33; *Bowen Investments Pty Ltd v Tabcorp Holdings Ltd* [2008] FCAFC 38, (2008) 166 FCR 494 [36], affd [2009] HCA 8, (2009) 236 CLR 272.

are not subject to the general law principles limiting compensatory damages, such as mitigation and remoteness.[66]

Substitutive damages are concerned to value the right infringed. There are various ways to assess this value. Three examples may be given.

To start, the value of the right may be determined by the market price of the right. For instance, if a defendant trespasser wrongfully evicts a claimant from several hotel apartments, and the claimant suffers minimal loss and the defendant makes no gain as a result of the wrong, the value of the proprietary right to the land is the reasonable license fee at the market rate.[67]

Next, the value of the right may be assessed by a 'difference in value' calculation. For example, in the case of non-delivery under a sale of goods contract, the value of a contractual right to buy goods at a particular price is the difference between the net contract price and the market value of the proprietary right which is the subject of sale, as at the breach date.[68]

Lastly, a court may employ the methodology of constructing a reasonable, hypothetical bargain. That is, the court creates a hypothetical bargain between a reasonable, willing seller and a reasonable, willing buyer for the seller to release the right in question.[69] The court may then adopt a figure equal to a proportion of the defendant's actual gain,[70] or the claimant's anticipated gain,[71] as a proxy for the value of the right. As the assessment is objective, the court works out what hypothetical, reasonable persons—not the parties themselves—would have agreed as a release fee.

II Compensation

Compensation is a loss-based response. It seeks to repair a loss, or to make good detriment. Accordingly, 'compensatory damages' refers to a financial remedy that seeks to repair a claimant's loss arising from a wrong.

---

66 Robert Stevens, 'Damages and the Right to Performance: a Golden Victory or Not?' in Jason W Neyers, Richard Bronaugh and Stephen GA Pitel (eds), *Exploring Contract Law* (Hart Publishing 2009) 181.

67 *Inverugie Investments Ltd v Hackett* [1995] 1 WLR 713 (PC).

68 *Williams Brothers v Ed T Agius Ltd* [1914] AC 510 (HL).

69 *Pell Frischmann Engineering v Bow Valley Iran Ltd* [2009] UKPC 45, [2011] 1 WLR 2370 [49].

70 *Experience Hendrix LLC v PPX Enterprises Inc* [2003] EWCA Civ 323, [2003] 1 All ER (Comm) 830 [46].

71 *Pell Frischmann Engineering v Bow Valley Iran Ltd* [2009] UKPC 45, [2011] 1 WLR 2370 [55]-[58].

Compensatory damages have four relevant features.[72] First, consequential losses are questions of fact. One must ask whether the claimant is *factually worse off* as a result of the wrong. Second, compensatory damages are assessed subjectively. That is, the claimant must prove the specific loss that she has suffered in order to recover. Third, they are assessed at the date of the judgment. Finally, compensatory damages are subject to several general law principles limiting recovery, such as mitigation and remoteness.

This thesis draws two important distinctions regarding loss. The first is temporal. This thesis juxtaposes 'anticipated' loss, which is the loss that a claimant potentially may suffer; and 'incurred' loss, which is the loss that a claimant in fact sustains. The second is qualitative. This thesis contrasts 'actual' loss, which is the prima facie factual loss that a claimant suffers; and 'legally-recoverable loss', which is the loss for which a claimant properly has a claim right after a court applies the general law damages rules (such as the mitigation and remoteness doctrines).

III Restitution and disgorgement

Restitution and disgorgement are gain-based responses.

As to the former, restitution seeks to reverse a transfer of value between the parties. Generally speaking, it means a defendant must 'give back' value he has obtained from a claimant. Where restitution responds to a wrong, the remedy is labelled 'restitutionary damages'. This remedy requires a defendant who has obtained a negative benefit by saving expense as a result of a wrong to pay to the claimant the equivalent of the amount saved. Why are restitutionary damages best interpreted in this way? In *A-G v Blake*, Lord Nicholls recognised that 'expense saved' is not within the scope of the disgorgement-based remedy of an account of profits, as ordinarily understood.[73] Accordingly, this definition will avoid an overlap[74] between the restitution and disgorgement responses to a wrong.

As to the latter, disgorgement seeks to divest or strip a gain. Broadly, it means a defendant must 'give up' value he has obtained. Accordingly, 'disgorgement damages' refers to a financial remedy that seeks to divest a defendant of his profits wrongfully made. Disgorgement damages have three relevant features.[75] First, consequential gains are questions

[72] Robert Stevens, *Torts and Rights* (OUP 2007) 60-61.

[73] [2001] 1 AC 268 (HL) 286.

[74] Text following n 84.

[75] Robert Stevens, *Torts and Rights* (OUP 2007) 61.

of fact. One must ask whether the defendant is *factually better off* by reason of the wrong. Second, disgorgement damages are assessed subjectively. That is, the claimant must prove the specific gain that the defendant has made in order for her to recover. Finally, they are assessed at the judgment date.

IV Perfection

A perfection response seeks substantially to achieve the very object of a causative event. Put another way, the purpose is factually to enforce one party's right to performance.[76] 'Performance' encompasses both positive acts and negative omissions.

An example: assume C and D enter into an agreement. D has a duty to perform any promises he has made, and C is able to enforce that duty through various remedies. Those remedies differ depending on the nature of the promise. If D promises to pay money, the action for the agreed sum enforces the duty. If D promises to do an act other than pay money, a specific performance order enforces the duty. However, if D promises to refrain from doing an act, then an (prohibitory) injunction enforces the duty. In short, these three remedies aim factually to 'perfect' the parties' consensual agreement.

V Punishment

The punishment response seeks financially to deprive a party where the deprivation is unrelated to the substituted value of a right, losses, gains, or achieving the object of a causative event. Punishment 'deprives' in the sense of leaving a party worse off—it places a party in a worse position to that which he or she occupied before doing an act.[77] There is no link between punishment and losses, or gains.[78] Where punishment responds to a wrong, the remedy is labelled 'punitive damages'. It is generally accepted that punishment is the essence of punitive damages.[79]

In *Cassell and Co Ltd v Broome*, Lord Reid described cases awarding punitive damages thus: 'The sums awarded as damages were more—sometimes much more—than could on any view be justified as compensatory, and courts … appeared to have permitted damages to be

---

[76] *Stern v McArthur* (1988) 165 CLR 489 (HCA) 522.

[77] Compare the disgorgement response. Requiring a party to disgorge his or her net gains simply places that party back in position he or she occupied prior to undertaking an act. Hence, the party is left no worse off.

[78] *James v Hill* [2004] NSWCA 301 [69].

[79] *Rookes v Barnard* [1964] AC 1129 (HL) 1221.

measured … by what the defendant ought to be made to pay as punishment for his outrageous conduct'.[80]

As discussed, a formalist approach attempts to take clues or derive insights about the private law's true nature from its form. Adopting a formalist mode of analysis, this passage suggests that civil punishment, as instantiated in punitive damages, contains two features. The first concerns the *conduct* punished. It is always the way in which the defendant breached a particular duty, rather than the breach itself. The second concerns the *structure* that the punishment takes. It is always a monetary payment parasitic upon, and excessively 'over and above', another payment made to instantiate a different response (such as compensation).

We must explore further the second feature. On the one hand, Australian courts have highlighted the parasitic nature of punitive damages. For instance, in *XL Petroleum (NSW) Pty Ltd v Caltex Oil (Australia) Pty Ltd*, Brennan J held that punitive damages 'are parasitic on compensatory damages …'.[81] On the other hand, jurists and judges have noted the excessive nature of punitive damages. As to the former, Beale states that courts award punitive damages 'as a punishment, so that the assessment goes beyond mere compensation of the claimant'.[82] As to the latter, Heydon JA has held that insofar as a defendant is 'required to pay more than could possibly be regarded as compensation he [is] being subjected to pure punishment'.[83]

Accordingly, a formalist analysis illustrates that the concept of civil punishment is one that is parasitic and excessive in structure: that is, a punitive response is consequent upon, but goes 'above and beyond', any substitutive, loss-based or gain-based response. As a result, in this thesis 'civil punishment' is used negatively: it refers to a financial response to an event that is provided simultaneously with another non-miscellaneous response, but cannot be assigned to any other category of non-miscellaneous response. Courts have labelled this notion a 'fine'. For example, in *Harris v Digital Pulse Pty Ltd*, Heydon JA held that 'the word "fine" is an entirely accurate description of that part of any award which goes beyond anything justly due to the [recipient] and is purely punitive'.[84]

D Comment

---

80 [1972] AC 1027 (HL) 1086.

81 (1985) 155 CLR 448 (HCA) 468-69.

82 Hugh Beale (gen ed), *Chitty on Contracts* (2nd supp, 31st edn, Sweet and Maxwell 2013) [26-043].

83 *Harris v Digital Pulse Pty Ltd* [2003] NSWCA 10, (2003) 56 NSWLR 298 [345].

84 [2003] NSWCA 10, (2003) 56 NSWLR 298 [345]. See also *Cassell and Co Ltd v Broome* [1972] AC 1027 (HL) 1100.

Two important features of the Birksian taxonomy must be noted.

First, a rationally defensible map must adhere to several rules. In particular, a classificatory scheme is undermined if there exists *gaps, overlaps* or *inconsistencies*. The Birkisan scheme adheres to these three requirements. To start, no gaps exist. The 'other' categories catch all miscellaneous events and responses falling outside the particularised categories. Although 'a sort of cheat',[85] this means that the Birksian taxonomy can never create gaps, or be incorrect. Next, the Birksian scheme contains categories of events and responses that are mutually exclusive, and hence it contains no overlaps. For example, compensation concerns the claimant's loss; disgorgement relates to the defendant's gain. Punishment, being parasitic in nature, concerns neither. The categorisations do not collide. Lastly, the Birksian taxonomy displays consistent categories selected according to the same criterion. So, the events-based categories all answer the question: what events generate rights? Further, the response-based categories all answer the question: what is the function of rights when they arise?

Second, some boxes in Table A have no content. For instance, the only response to unjust enrichment is restitution; the other responses fail to arise. Accordingly, the box at the intersection of unjust enrichment and compensation, or unjust enrichment and punishment, has no content.[86] This is unproblematic, for the Birksian grid—as originally devised or as extended—is intended simply to identify possibilities, and a causative event's logic or rationale may demand that one response be available, but preclude another. This classification scheme is intended to highlight the need to *consider* every potential response to each events-based category. However, 'At the same time that it opens new avenues of relief, Birks's taxonomy also helps to eliminate responses that ought not to exist'.[87]

6 Conclusion

This chapter's goal has been to present an overview of this thesis, outline its scope, and address key terminological, methodological and classificatory issues. It has done so in some detail because a developed understanding of these matters is critical for the analysis that follows.

---

85 Peter Birks, 'Equity in the Modern Law: An Exercise in Taxonomy' (1996) 26 UWAL Rev 1, 9.

86 Peter Birks, *Unjust Enrichment* (2nd edn, OUP 2005) 25.

87 Mitchell McInnes, 'Taxonomic Lessons for the Supreme Court of Canada' in Charles Rickett and Ross Grantham (eds), *Structure and Justification in Private Law* (Hart Publishing 2008) 89.

Part One: Restatement

2 HISTORY

This chapter begins the process of restating the Penalty Rule. It briefly analyses the rule's origins and historical development. It does so in five steps. First, it comments on the particular strategy this thesis adopts in exploring historical issues. Second, the chapter considers the rule's origins. Third, it outlines the early legislative developments. Fourth, the chapter examines a time when the parties' autonomy reigned supreme. Finally, it identifies the key developments that grounded Lord Dunedin's judgment in *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd*,[1] the most 'central' case of all in this area.

1 Strategy

This chapter is very short. There is good reason for this. It necessarily follows from the methodological approach that this thesis adopts; the inherent nature of any historical analysis; and the Penalty Rule's peculiar historical development. Each point is considered in order.

First, as Chapter 1 explained, this thesis adopts an interpretive methodology. An interpretive account of an area of law aims to enhance legal understanding (by revealing an intelligible order in the law and highlighting its meaning), and to provide the 'best' explanation of legal data. Contrast an historical account, which seeks to explain how and why the legal area has developed in the way it has. An interpretive account will *draw on* an historical account, but the two are distinct. Any historical analysis can only provide helpful hints as to how we should best understand the current law. As Smith argues, 'History only provides clues to the answers that the interpreter seeks. What an interpreter is trying to understand—to interpret—is the law, not what motivated the judges who made the law.'[2] Accordingly, any historical argument can take an interpretivist only so far. An interpretive analysis, then, should elucidate historical points only to the extent necessary to highlight the contemporary law's order and meaning, and to the degree needed best to explain that law. This chapter takes precisely that approach.

---

1 [1915] AC 79 (HL).

2 Stephen A Smith, *Contract Theory* (OUP 2004) 6.

Second, history is important, but it is not justification. Given the Penalty Rule's longevity, one might seek to explain the rule adopting a purely historical analysis. Unquestionably, the result of that analysis would be edifying. But it would not be justificatory. As Beever explains, as an historical investigation 'explains attitudes such as anti-Semitism without defending them, it explains the existence of legal rules without supporting them. But this is never enough in law. You cannot justify a legal rule by elucidating how it came to be.'[3] Consequently, an historical analysis will not suffice to satisfy any aim of this thesis set out in Chapter 1. Hence, it is appropriate to undertake an historical analysis, but only to the extent that it assists with an interpretive approach.

Finally, antiquated decisions are generally an unreliable basis upon which to determine the present law. This applies with redoubled force to the Penalty Rule.

It is well known that the rule pursued 'a tortuous path in the course of its long development'.[4] For example, many of the historical cases are contradictory. Thus, McGregor argues that 'the numerous 19th century cases show some confusion and not infrequent difficulty in reconciling, a factor which must be recognised when relying on them as precedents'.[5] It follows, then, that it is 'a risky enterprise' to construct an argument based on the old decisions.[6] The exercise is risky because its product it is inherently unreliable, as a matter of strict doctrine.

Given its difficult history, there exists a 'good deal of disagreement' as to how the rule evolved.[7] Many examples exist. Two will be given. For one, in *AMEV-UDC Finance Ltd v Austin*, Mason and Wilson JJ examined the rule's history to conclude that A was not entitled partially to enforce a penalty clause against B, whereas Deane J conducted a similar, careful historical analysis to arrive at the exact opposite conclusion.[8] For another, in *AMEV*, Mason and Wilson JJ concluded that, although historically equity courts relieved in penal bond cases where no express promise to perform the condition existed, in many cases a promise could 'readily be implied'.[9] Conversely, in *Andrews v Australia and New Zealand Banking Group Ltd*,

---

[3] Allan Beever, *The Law of Private Nuisance* (Hart Publishing 2013) 6.

[4] *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 186.

[5] Harvey McGregor, *McGregor on Damages* (19th edn, Sweet and Maxwell 2014) [15-007] (footnote omitted).

[6] *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 186.

[7] *Philips Hong Kong Ltd v A-G of Hong Kong* (1993) 61 BLR 41 (PC) 55.

[8] (1986) 162 CLR 170 (HCA) 191 cf 203, 205.

[9] *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 190.

after its historical review the Australian High Court flatly rejected this conclusion.[10] In short, if one wishes to mount an historical argument, one probably can find a case or a writer to support most propositions.

The short point is this: any extended historical analysis of the Penalty Rule likely will be more time-consuming than useful.

2 Origins

By most accounts, the Penalty Rule originated in equity. It derives from the Court of Chancery's practice of granting relief from penal bonds. A brief outline of that practice follows.

A penal bond took the following form. X promised to pay Y a stated sum of money on a specified date. X's promise was subject to a condition that if a particular event occurred or did not occur on or before a certain date, X's promise to pay would be void. Thus, if the condition was met, X's obligation to pay was released. If the condition was unsatisfied, X's promise to pay became unconditional. Y was then entitled to recover the stated sum as a debt. Two forms of penal bond existed. The first was a 'money bond'. Under this form of bond, the condition involved X paying a fixed sum. The fixed sum may have represented a loan amount that X was to repay to Y, or the price of goods or services that X had bought from Y. The second was a 'performance bond'. Pursuant to this form of bond, the condition concerned X performing a particular act other than paying money.

Originally, common law courts strictly enforced penal bonds. A judge at law would only discharge a bond if the obligor had precisely complied with the condition. The approach was rigid. As such, at times it resulted in 'harsh' outcomes.[11]

In the 15th century, equity courts began granting relief by restraining the claim brought for the penalty. This jurisdiction was originally very narrow. Essentially, it only covered cases of clear hardship. Two examples will suffice.[12] Thus, an equity judge granted relief where the obligor had paid in full the sum due under the bond, but had not obtained a sealed release.

---

[10] [2012] HCA 30, (2012) 247 CLR 205 [66].

[11] *Andrews v Australian and New Zealand Banking Group Ltd* [2011] FCA 1376, (2011) 211 FCR 53 [10], revd [2012] HCA 30, (2012) 247 CLR 205.

[12] AWB Simpson, 'The Penal Bond with Conditional Defeasance' (1966) 82 LQR 392, 416; Chris Rossiter, *Penalties and Forfeiture* (Lawbook Company 1992) 5-6.

Further, the Court of Chancery intervened where the bond appeared to be unconditional, but in fact was conditional, and the obligor had satisfied the condition.

By the end of the 16th century, the jurisdictional scope widened to cover cases involving fraud, accident, mistake and surprise, and new cases of hardship. For instance, if an obligor had paid the amount due in order to fulfil the condition, but had paid late due to 'accident, mistake or extremity',[13] an equity judge would issue an injunction restraining the obligee from suing on the bond at common law.

In the early 17th century, the Court of Chancery's power to relieve against penal bonds broadened. It became unnecessary for the obligor to establish a specific ground of relief. Rather, the following regime was established. An equity court relieved the obligor from performing the (main) penal obligation. It did so by enjoining the obligee from enforcing her bond at common law, or restraining the obligee from executing her judgment obtained at law. However, the court would also order that the obligor make good the obligee's loss arising from the event contained in the (subsidiary) condition occurring, or failing to occur.[14] A simple example: suppose X agreed to pay to Y a penal money bond of £40 subject to a condition that if X repaid to Y a loan amount of £5 by the due date, the bond would be void. An equity judge would relieve against the penal money bond by enjoining Y from enforcing her bond at law, but ordering that X repay to Y the original debt of £5, with interest and costs.

There are three important points to note.

To start, equity courts relieved against penal bonds exacting a penalty upon the occurring, or not occurring, of an event. That event, however, did not necessarily involve a breach of contract. To explain, under a penal bond arrangement, the obligor's promise to pay the penal sum constituted an obligation. As stated, the obligor's obligation was subject to a condition. This condition need not have been promissory in nature. In many cases, the obligor did not promise to ensure that the event contained in the condition occurred, or did not occur. Put differently, the obligor was not under any *duty* regarding the condition. Accordingly, technically the obligor did not breach any obligation when the event failed to occur, or in fact occurred.[15] But the Court of Chancery granted relief nonetheless.

---

[13] *Andrews v Australian and New Zealand Banking Group Ltd* [2011] FCA 1376, (2011) 211 FCR 53 [11], revd [2012] HCA 30, (2012) 247 CLR 205.

[14] *Jobson v Johnson* [1989] 1 WLR 1026 (CA) 1039; *Andrews v Australian and New Zealand Banking Group Ltd* [2011] FCA 1376, (2011) 211 FCR 53 [12]-[13], revd [2012] HCA 30, (2012) 247 CLR 205.

[15] *Bridge v Campbell Discount Co Ltd* [1962] AC 600 (HL) 629-31; *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 197-98.

Next, in some cases, the obligor brought a claim after having already paid the penal sum to the obligee. If so, an equity court would require the obligee to repay to the obligor the excess sum. In penal money bond cases, this comprised the sum over an amount equivalent to the principal sum, with interest and costs.[16]

Lastly, the basis for judicial intervention evolved during this period. At first, the Court of Chancery was motivated to mitigate the obvious hardship that the common law courts caused through strictly enforcing penal bonds. Later, equity judges also became concerned with alleviating additional forms of hardship, and redressing vitiated consent. According to Simpson, by the 17th century, the courts recognised that 'the general aim of the law was to compensate rather than to punish', such that a party 'should only be permitted to recover *compensation* for loss actually suffered through default'.[17] Thus, it is said that equity courts developed a concern to prevent 'overcompensation'.[18]

3 Legislation

By the mid-17th century, the Court of Chancery relieved against penal bonds as a matter of course. Further, it also began to relieve against penalty clauses in simple contracts. Simultaneously, the common law courts significantly changed their procedures in actions brought in assumpsit. As a result, courts of law began granting relief which emulated the relief that the Chancery Courts gave, at least in the case of penal money bonds.[19] These developments saved obligors and contracting parties the time and expense of obtaining an equitable injunction to restrain any action at law.

English Parliament then enacted two statutes.

The first statute was known as the Statute of William III. It dealt with penal performance bonds. In summary, the statute provided as follows.[20] If the plaintiff instituted proceedings at law to recover a penal sum due on a bond, or due as result of a breach of covenant, upon proving a breach of condition or covenant the plaintiff was entitled to judgment for the full amount. However, the plaintiff was required to plead any alleged breach

---

16 *Friend v Burgh* (1679) Fin 437, 23 ER 238.

17 AWB Simpson, 'The Penal Bond with Conditional Defeasance' (1966) 82 LQR 392, 419-20 (original emphasis).

18 Kanaga Dharmananda and Leon Firios, 'Penalties Arising Without Breach: The Australian Apogee of Orthodoxy' [2013] LMCLQ 145, 147.

19 *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [53], [55]; Sarah Worthington, *Equity* (2nd edn, OUP 2006) 224.

20 Administration of Justice Act 1696 (8 & 9 Will 3 c 11) s 8.

or breaches. A jury then assessed the plaintiff's incurred loss. Following this, if the defendant paid into court the damages as assessed (with interest and costs), the court ordered a stay of execution.

Notably, under the statute's provisions, 'The concept of breach was central.'[21] Further, the penal sum set the maximum amount that a plaintiff might recover from the defendant upon the bond: the obligor was 'not answerable in the whole beyond the amount of the penalty'.[22]

The second statute was known as the Statute of Anne. It concerned penal money bonds. The statutory regime was this.[23] If the plaintiff brought an action in debt on a bond at law, the plaintiff was entitled to judgment. However, if the defendant paid into court the principal amount owing (with interest and costs), the court held that the defendant had satisfied all liability under the bond.

The statutes provided a plaintiff with a clear basis for obtaining relief from a penalty in a court of law. Despite this, during the 18th century, many plaintiffs chose to sue in Chancery. Hence, the equitable jurisdiction continued to develop. Further, although in many cases the statutes seemingly applied, the common law courts generally ignored the statutes. However, as the legislative approaches mirrored the general law jurisdiction,[24] the courts reached outcomes consistent with the results that the statutes demanded.[25]

4 Autonomy

During the 18th century, the penal bond metamorphosed into a clause stating that X was to pay to Y a fixed sum of money upon a breach of contract. This turned the penal bond on its head: now the obligation to pay the penal sum was the subsidiary obligation, and not the main obligation.

At this point, common law and equity courts prohibited a party enforcing a penalty clause. However, they enabled a party to enforce a liquidated damages clause. The Statute of William III and Statute of Anne encapsulated the same approach, in that the statues applied to

---

[21] *Andrews v Australian and New Zealand Banking Group Ltd* [2011] FCA 1376, (2011) 211 FCR 53 [24], revd [2012] HCA 30, (2012) 247 CLR 205.

[22] *Wilde v Clarkson* (1795) 6 D & E 303, 304; 101 ER 566, 566.

[23] Administration of Justice Act 1705 (4 & 5 Anne c 3) ss 12-13.

[24] AWB Simpson, 'The Penal Bond with Conditional Defeasance' (1966) 82 LQR 392, 419.

[25] eg *Sparrow v Paris* (1862) 7 H & N 594, 599; 158 ER 608, 610.

penal sums, but did not apply to liquidated damages. Hence, the Acts merely emphasised the already present need to distinguish between liquidated damages and penal sums.[26]

From the end of 18th century until the modern law was stated in *Dunlop*, courts adopted the 'intention test' to discriminate between a liquidated damages clause and a penalty clause. Under this test, courts differentiated between the two types of clauses based on the parties' intention. Judges gleaned that intention from the label the parties used to describe the sum.[27] Accordingly, if the contract labelled the relevant amount as 'liquidated damages', the court concluded that the parties had intended the payment be in the nature of liquidated damages, so the entire sum was recoverable.[28] Conversely, if the contract referred to the amount as a 'penalty', the court would invalidate the clause for that reason alone.[29]

Two aspects of the court's practice in the 18th century must be highlighted.

On the one hand, if the court held that the agreed damages clause was valid, A recovered the liquidated damages from B at law 'as a debt'.[30]

On the other hand, if the court held that the clause was invalid, A maintained a right of election. A could choose to sue B on the penalty clause in debt, or claim from B general damages in assumpsit.[31] However, whether A sued in debt or in assumpsit, the result was generally the same: A recovered from B an amount equal to her incurred loss.

5 Precursors

Throughout the 19th century, courts continually held that the intention test determined whether an agreed damages clause was valid, or not. However, courts also began to qualify the test. The parties' intention as encapsulated in the label remained of 'paramount importance'.[32] However, occasionally judges were prepared to look beyond the parties' intention as expressed in the label. Several qualifications were recognised. Two were as follows.

---

26 See *Davies v Penton* (1827) 6 B & C 216, 224-25; 108 ER 433, 436.

27 *Wallis v Smith* (1882) 21 Ch D 243 (CA) 266; *Waterside Workers' Federation of Australia v Stewart* (1919) 27 CLR 119 (HCA) 128.

28 *Reilly v Jones* (1823) 1 Bing 302, 305-306; 130 ER 122, 123.

29 *Smith v Dickenson* (1804) 3 Bos & P 630, 632; 127 ER 339, 341.

30 Chris Rossiter, *Penalties and Forfeiture* (Lawbook Company 1992) 13.

31 *Beckham v Drake* (1849) 2 HL Cas 579, 615, 645; 9 ER 1213, 1226, 1237.

32 Chris Rossiter, *Penalties and Forfeiture* (Lawbook Company 1992) 20.

The first qualification was that, in some cases, the courts *imputed* to the parties an intention to characterise the sum as penal. The courts did so irrespective of the parties' actual intention, as manifested in the contract's express terms. Two examples will suffice.

For one, at the time, general compensatory damages were unrecoverable for the late payment of a debt. Consequently, where a clause imposed upon B an obligation to pay a larger sum to A consequent upon B's failing to discharge his duty to pay a smaller sum, the courts inferred that 'the parties considered the larger sum [to be] a penalty'.[33] Thus, in *Kemble v Farren*, the parties had described the relevant amount as 'liquidated and ascertained damages',[34] but the Court nonetheless held that the sum was penal because the clause fell within this qualification. Tindal CJ reasoned that if 'a very large sum should become immediately payable … in consequence of the non-payment of a very small sum',[35] then the parties must have considered that the former obligation was penal.

For another, in some cases, the agreed damages clause provided that B was to pay a single lump sum to A allegedly as compensation upon one or more events occurring, some of which might cause serious but others only trifling loss. In such cases, it was said that the court was 'not influenced solely by the words used'; rather, the court looked at the parties' intention.[36] As a result, where the clause was indiscriminate in nature, the courts presumed that 'the parties intended the sum to be penal, and subject to modification',[37] despite its label as liquidated damages. Hence, in *Sandeman v Bedwell*, although the contract referred to the stated sum as 'liquidated damages',[38] the New South Wales Supreme Court invalidated the clause on the basis that the relevant presumption was sustained.[39] Sir W Manning J held that where the agreed damages sum secured 'the performance of things of different degrees of importance, the Court [would] hold that it could not have been the intention of the parties, whatever they [had] said, to exact [liquidated damages]'.[40]

The second qualification was that, on occasion, courts suggested that they were prepared to *disregard* the parties' expressed intention and impugn a clause as penal. For

---

[33] *Astley v Weldon* (1801) 2 Bos & P 346, 353; 126 ER 1318, 1323.

[34] (1829) 6 Bing 141, 147; 130 ER 1234, 1236.

[35] *Kemble v Farren* (1829) 6 Bing 141, 148; 130 ER 1234, 1237.

[36] *Gleeson v Kingston* (1880) 6 VLR (L) 243 (VSCFC) 246.

[37] *Lord Elphinstone v Monkland Iron and Coal Co Ltd* (1886) 11 App Cas 332 (HL) 342.

[38] (1879) 2 SCR (NS) (NSW) 145 (NSWSC) 145.

[39] See similarly *Lord Elphinstone v Monkland Iron and Coal Co Ltd* (1886) 11 App Cas 332 (HL).

[40] *Sandeman v Bedwell* (1879) 2 SCR (NS) (NSW) 145 (NSWSC) 147.

instance, some judges intimated that they would be willing to cast aside the parties' intention as displayed in the label if the clause was 'oppressive' in nature;[41] if the agreement involved a 'manifest injustice';[42] or if there existed a 'gross' inequality between the parties.[43]

In the early 20th century, courts commonly recognised many instances of imputed and disregarded intention. As the number of exceptional cases grew, the importance of the intention test was slowly eroded. Thus, judges began to accept that a different principle applied in determining whether an agreed damages clause was valid or not: that is, the parties' use of the expression 'penalty' or 'liquidated damages' was relevant, but not conclusive. The court's overriding mandate was to discern whether the payment was *in truth* a liquidated damages sum, or a penal sum.

At this point, we must note five issues important to our understanding of the law.

First, the key developments outlined in this section formed the basis of Lord Dunedin's judgment in *Dunlop*, discussed in Chapter 3. That case has been described as a '*locus classicus*'.[44] In essence, *Dunlop* is to the law of penalties what *Donoghue v Stevenson*[45] is to the law of negligence.

Second, at this juncture in the Penalty Rule's history, the 'core case' for engaging the rule involved a breach of contractual obligation.[46]

Third, at this stage in the rule's historical development, 'no clear doctrinal basis was stated for the court's power of intervention'.[47]

Fourth, at this time, the rule's theoretical rationale was seemingly to prevent 'overcompensation'. So, Rossiter argues that the 'historical role' of the Anglo-Australian rule was 'to minimise over-compensation to [A] and to ensure that [B was] only rendered liable for the true loss' that A sustained.[48]

---

[41] *Thompson v Hudson* (1869-70) 4 LRHL 1 (HL) 30.

[42] *Gleeson v Kingston* (1880) 6 VLR (L) 243 (VSCFC) 246.

[43] *Astley v Weldon* (1801) 2 Bos & P 346, 351; 126 ER 1318, 1321.

[44] *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [55].

[45] [1932] AC 562 (HL).

[46] eg *Kemble v Farren* (1829) 6 Bing 141, 148; 130 ER 1234, 1237; *Reynolds v Bridge* (1856) 6 E & B 528, 541; 119 ER 961, 966.

[47] DW Greig and JLR Davis, *The Law of Contract* (Lawbook Company 1987) 1447.

[48] Chris Rossiter, *Penalties and Forfeiture* (Lawbook Company 1992) 35.

Finally, Rossiter notes that historically the private law recognised that A was entitled to 'just' compensation, being 'compensation known to the law, not necessarily compensation which the parties thought fair'.[49] Thus, the writer implies that the preventing 'overcompensation' rationale was premised upon some overriding, mandatory norm that dictated the amount of compensation that A was able to recover.

6 Conclusion

The Penalty Rule has a long, complicated and difficult history. This chapter has provided a very brief outline of the rule's development over the centuries. It has done so not to provide an account of the doctrine's causal history, but merely to highlight as much of that history as is necessary for one fully to understand—and interpret—the modern law.

[49] Chris Rossiter, *Penalties and Forfeiture* (Lawbook Company 1992) 35.

3 ORTHODOXY

One must clearly identify any private law rule before seeking properly to interpret it. Accordingly, this chapter examines the orthodox view of the modern Penalty Rule. It first outlines the discrete stages of any analytical inquiry under the rule. The chapter illustrates that, as conventionally understood, the rule requires that a court address three issues. The chapter is then organised thematically according to these discrete issues. The discussion of each stage of the penalty inquiry lays the foundation for the chapters that follow.

1 Three stages

The Penalty Rule is difficult. In order to understand it, one must clearly identify the structure of the rule, and distinguish the discrete stages of the analytical inquiry that it requires. Without doing so, it is nearly impossible to make sense of the case law.

Historically, courts have failed rigorously to distinguish the separate stages of the analytical process. A recent example is *Andrews v Australia and New Zealand Banking Group Ltd*, where the Australian High Court defined a penalty clause and the Penalty Rule in the following terms:

> In general terms, a stipulation prima facie imposes a penalty on a party (the first party) if, as a matter of substance, it is collateral … to a primary stipulation in favour of a second party and this collateral stipulation, upon the failure of the primary stipulation, imposes upon the first party an additional detriment, the penalty, to the benefit of the second party … If compensation can be made to the second party for the prejudice suffered by failure of the primary stipulation, the collateral stipulation and the penalty are enforced only to the extent of that compensation …[1]

In this passage, the High Court refers to the three stages of the analytical inquiry, but does not clearly identify them. First, it states that a penalty clause is a collateral stipulation that imposes a penalty upon the primary stipulation failing. Thus, a court must first enquire into whether the clause is properly an agreed damages clause, such that the court possesses the *jurisdiction* to review the clause under the Penalty Rule. Second, the Court notes that a penalty constitutes 'an additional detriment' placed upon B. In other words, a court must next examine whether the agreed damages clauses passes or fails the relevant test or tests of *validity*. Finally, the

[1] [2012] HCA 30, (2012) 247 CLR 205 [10].

judgment notes that a court will enforce a penalty clause 'only to the extent of [any relevant] compensation'. So, if the agreed damages clause fails the test or tests of validity, a court must declare the clause penal, and then determine the *consequences* of invalidity.

Accordingly, the structure of the Penalty Rule requires a three-stage approach: jurisdiction, validity and consequences. The remainder of this chapter examines the traditional understanding of each distinct stage of the penalty inquiry.

2 Jurisdiction

A court can only review an agreed damages clause. The orthodox view is that a clause is an agreed damages clause only if it provides for B to transfer a benefit to A upon B's breach of his contractual duty owed to A. The case best illustrating this principle is the 'central' case of *Export Credits Guarantee Department v Universal Oil Products Co.*[2]

The facts were complicated. For present purposes, a simplified account suffices. Procon contracted to construct a refinery for Provincial. Export Credits financed the project, acting as guarantor. Procon agreed with Export Credits that, if the guarantee was called upon at a time when Procon was in breach of its contract with Provincial, Procon would pay Export Credits a sum equivalent to that paid out under the guarantee. Export Credits paid out £39 million under the guarantee. At the time, Procon was in breach of its contract with Provincial. Export Credits demanded that Procon reimburse it. Procon claimed the relevant clause was penal.

The case raised the following issue: did the Penalty Rule apply to a clause that required Procon (B) to pay a stated sum to Export Credits (A) if B breached one or more of its duties owed to a third party, Provincial (C)?

Agreeing with the courts below, the House of Lords held the clause was unreviewable. Lord Roskill reasoned that the rule did not apply because the clause 'provided for payment of money on the happening of a specified event other than a breach of contractual duty owed by the contemplated payer to the contemplated payee'.[3] In short, that B has breached *a* contractual duty does not suffice. B's breach of duty must constitute an infringement of A's right before the court may review the clause.

3 Validity

[2] [1983] 1 WLR 399 (HL).

[3] *Export Credits Guarantee Department v Universal Oil Products Co* [1983] 1 WLR 399 (HL) 402.

The tests of validity that traditionally apply under the Penalty Rule derive from Lord Dunedin's judgment in the 'central' case of *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd*.[4] Courts and commentators have variously described these tests as 'authoritative',[5] and as rules upon which 'the modern law rests'.[6] Anglo-Australian appellate courts have applied these tests of validity countless times.

The facts were straightforward. Dunlop manufactured motor tyres, covers and tubes. It supplied these to its dealer, New Garage, on terms. New Garage agreed not to do any of the following: tamper with Dunlop's markings on the covers or tubes; sell or offer the goods to customers at prices below Dunlop's current list prices; sell the goods to persons whose supplies Dunlop had decided to suspend; or exhibit or export the goods without Dunlop's consent. Under clause 5, New Garage agreed to pay to Dunlop £5 as 'liquidated damages' for each product sold or offered in breach of the agreement. New Garage sold covers and tubes at prices below Dunlop's set prices. Dunlop sued. In its defence, New Garage argued that clause 5 was penal.

During his speech, Lord Dunedin attempted to synthesise the various approaches the courts took to the validity issue in the 19th century and early 20th century, some of which approaches were discussed in Chapter 2. Thus, his Lordship listed several propositions for identifying a penalty clause, as follows.[7]

First, the parties' use of the phrase 'penalty' or 'liquidated damages' is relevant but not conclusive. In short, labels are not determinative.

Second, the essence of a penalty is a sum payable *in terrorem* of the breaching party, B; the essence of liquidated damages is a genuine pre-estimate of A's loss. In this thesis, the 'Estimation Test' refers to the test that determines whether an agreed damages clause is valid based on whether it represents a genuine pre-estimate of loss.

Finally, in construing the contract, the following four tests 'may prove helpful, or even conclusive'. The first is the 'Extravagance Test'. It holds that a sum is penal if it is 'extravagant and unconscionable in amount' compared to 'the greatest loss that could … have followed from the breach'. The second test is the 'Greater Sum Test'. It states that a clause is

---

[4] [1915] AC 79 (HL).

[5] *Murray v Leisureplay plc* [2005] EWCA Civ 963, [2005] IRLR 946 [44].

[6] Harvey McGregor, *McGregor on Damages* (19th edn, Sweet and Maxwell 2014) [15-008].

[7] *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd* [1915] AC 79 (HL) 86-88.

penal if the breach consists only in not paying a sum, and the sum stipulated in the clause is a greater sum. The Greater Sum Test, it was said, is 'truly a corollary' to the Extravagance Test. The third is the 'Indiscriminate Cause Test'. It holds that a presumption of penalty arises when a single lump sum is payable upon one or more of several events occurring, some of which may cause serious harm and others trivial harm. The final test is the 'Impossibility Test'. Under it, the fact 'precise pre-estimation is almost an impossibility' will not render a clause penal; on the contrary, in such circumstances, it is more likely that the clause will be upheld.

His Lordship added two further points. For one, the true nature of the agreed damages clause is decided 'upon the terms and inherent circumstances of each particular contract …'. For another, the tests of validity are applied as at the contract date, not a later date (such as the date of the breach, or of the judgment). That is, the tests are applied *ex ante*, not *ex post*.

On the facts, the House of Lords unanimously held that clause 5 was valid. Their Lordships were willing to assume that clause 5 applied to several of New Garage's obligations, not merely the obligation not to sell or offer the goods below the list prices. Therefore, the House of Lords held that the term failed the Indiscriminate Cause Test. This raised a presumption of penalty. However, the presumption was rebutted: the loss likely to result from any breach was 'of such an uncertain nature'[8] that it could not be accurately assessed. One aspect of the anticipated loss that was apparently impossible to assess was the undermining of Dunlop's agency system of marketing and distribution. In Lord Dunedin's view, the undercutting of the integrity of Dunlop's wholesale distribution system constituted an 'indirect' loss which was 'impossible to forecast'.[9]

4 Consequences

Conventionally speaking, if a court declares an agreed damages clause valid, A may enforce the liquidated damages clause against B according to its terms. However, if the court finds that the clause is invalid, one of two consequences may follow. On the one hand, A may sue B for general damages for breach of contract. On the other hand, the penalty clause may be 'compressed' to reflect A's incurred loss, and A may sue B on the clause to recover the amount of that loss. Traditionally speaking, an English claimant may elect between these options; an Australian claimant may only sue for general damages.

[8] *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd* [1915] AC 79 (HL) 96.

[9] *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd* [1915] AC 79 (HL) 88.

A 'central' Australian case illustrating the first option is *AMEV-UDC Finance Ltd v Austin*.[10] The parties entered into a contract for the hire purchase of printing equipment. The contract provided that if the hirer defaulted in paying a rental instalment, the owner obtained the right to terminate the contract and repossess and resell the equipment. The owner would also become entitled to claim under an agreed damages clause in the agreement. That clause calculated the agreed damages amount by reference to the owner's 'lost bargain': namely, any unpaid rental arrears, the balance of the rental instalments for the unexpired term of the agreement, the difference in value between the equipment's specified residual value and the anticipated resale price, and interest on the total. The hirer then defaulted, which act amounted to a non-repudiatory[11] breach. The owner exercised its contractual right to terminate the contract and sought to enforce the agreed damages clause.

Before the Australian High Court the owner conceded that the agreed damages clause was penal. The clause was penal because it failed to provide for two factors: first, a rebate for the accelerated payment of future instalments; and, second, a credit for the anticipated sale proceeds if they exceeded the residual value of the goods.[12] In other words, the agreed damages clause did not allow for A to recover from B the total rental less the excess of the anticipated resale price over the residual value of the goods, nor did it discount for early receipt.

Accordingly, the Court only needed to decide the appropriate consequential relief. However, the judges sharply divided on the proper consequences of an agreed damages clause being declared penal. The majority held that A is only entitled to claim against B general damages for breach of contract. Mason and Wilson JJ held that 'Without exception the cases show that once the agreed sum is held to be a penalty the [claimant] recovers damages for breach of contract in lieu of the penalty'.[13] In dissent, Deane J held that A is entitled partially to enforce a penalty clause up to the amount of A's actual incurred loss.[14]

---

[10] (1986) 162 CLR 170 (HCA).

[11] A 'repudiatory breach' includes a breach of condition, a sufficiently serious breach of an intermediate term and a repudiation: see *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 186; *Hongkong Fir Shipping Co Ltd v Kawasaki Kisen Kaisha* [1962] 2 QB 26 (CA) 69-72. If a defendant commits a repudiatory breach, a claimant obtains the right to terminate at law and claim 'bargain damages': that is, damages equivalent to the full value of the lost performance.

[12] *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 180-81, 209-10.

[13] *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 191. See similarly 175, 176 (Gibbs CJ).

[14] *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 203, 205.

On the facts, a bare majority of the High Court[15] held that the owner was only entitled to recover the loss flowing directly from the hirer's breach, namely, the outstanding rental arrears with interest.[16] However, the owner was *not* entitled to recover the additional loss, namely the present value of the future instalments as at the termination date. Under orthodox causation rules, any loss flowing from a termination based on a non-repudiatory breach is seen as resulting from a party's own act in electing to terminate, not any anterior breach.[17] Accordingly, the additional loss resulted from the owner's election to terminate the agreement, not the hirer's breach of it.[18] Therefore, it was unrecoverable.

A 'central' English case illustrating the second option is *Jobson v Johnson.*[19] The (simplified) facts were these. Under a sale of shares contract the purchaser agreed to buy from the vendor the proprietary rights to several shares in a football club. The contract price was £351,698. The contract provided that the purchaser was to pay the price by an initial instalment of £40,000, followed by six half-yearly instalments of £51,948. The agreement contained a clause stating that, if the purchaser defaulted in paying the second or any subsequent instalment, he would retransfer to the vendor the proprietary rights to the shares for £40,000. The vendor transferred the rights to the shares. The purchaser paid £140,000 toward the price. He then defaulted. The vendor sought specific performance of the repurchase clause.

The English Court of Appeal unanimously held that the retransfer clause was a penalty clause. Relevantly, as to the consequences of invalidity, the Court held that penalty clauses are partially enforceable, such that A may sue B on the penalty clause. In Nicholls LJ's words:

> the strict legal position is not that [a penalty] clause is simply struck out of the contract, as though with a blue pencil, so that the contract takes effect as if it had never been included therein. Strictly, the legal position is that the clause remains in the contract and can be sued upon, but it will not be enforced by the court beyond the sum which represents … the [incurred] loss of the party seeking payment.[20]

---

[15] Gibbs CJ, Mason and Wilson JJ, Deane J and Dawson J dissenting.

[16] The owner was in principle entitled to recover the amount of any deficiency on resale, but this was irrelevant on the facts.

[17] *ENE 1 Kos Ltd v Petroleo Brasileiro SA Petrobras (The Kos)* [2010] EWCA Civ 772, [2011] 2 All ER (Comm) 57 [16], revd [2012] UKSC 17, [2012] 2 AC 164.

[18] *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 175, 176, 186, 191.

[19] [1989] 1 WLR 1026 (CA).

[20] *Jobson v Johnson* [1989] 1 WLR 1026 (CA) 1040.

If a clause is penal, 'to that extent it is unenforceable. But it is enforceable to the extent that it is not a penalty …'.[21] So, if a penalty clause exists, the court will engage in a 'scaling down' exercise: it will relieve by 'cutting down the extent to which the contractual obligation is enforceable'.[22] The court scales down the penal amount to a sum that equals the innocent party's incurred loss. In short, the clause is 'compressed' to reflect A's incurred loss, and A may then sue B upon the clause to recover.

On the facts, the majority of the Court of Appeal[23] gave the vendor two options.[24] The vendor could elect for the court to sell the rights to the shares, and have his incurred loss—the unpaid instalments and interest—paid from the sale proceeds. Alternatively, the vendor could opt for a specific performance order requiring the retransfer of the rights to the shares for £40,000 if, on inquiry, the net value of the bargain did not exceed the amount of the vendor's incurred loss: that is, the value of the shares did not exceed by more than £40,000 the total amount of the unpaid instalments plus interest. According to Nicholls LJ, these options were 'the equivalent [on the facts] to the automatic scaling down of a (pecuniary) penalty clause'.[25] The relief merely amounted to invalidating the clause's penal element. As Peel notes, 'in this respect it resembled the legal consequence of a penalty clause in the normal sense of that expression'.[26]

5 Conclusion

This chapter has illustrated that the structure of the Penalty Rule requires a three-stage analytical approach. One must consider the matter of jurisdiction, questions of validity, and consequential issues. The chapter then outlined the orthodox view of how the rule operates at each distinct stage, based on the central cases.

Chapters 4 to 6 analyse the traditional view of the rule's jurisdictional test, the tests of validity and the consequences of applying the rule. They seek to build on the approaches found in the cases examined in this chapter, and to provide the best reconstruction of those approaches.

---

[21] *Jobson v Johnson* [1989] 1 WLR 1026 (CA) 1049. See similarly *Murray v Leisureplay plc* [2005] EWCA Civ 963, [2005] IRLR 946 [29].

[22] *Jobson v Johnson* [1989] 1 WLR 1026 (CA) 1042. See also *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [126].

[23] Dillon LJ and Nicholls LJ, Kerr LJ dissenting.

[24] *Jobson v Johnson* [1989] 1 WLR 1026 (CA) 1037, 1045.

[25] *Jobson v Johnson* [1989] 1 WLR 1026 (CA) 1045.

[26] Edwin Peel, *The Law of Contract* (13th edn, Sweet and Maxwell 2011) [20-153].

4 JURISDICTION

As traditionally understood, an agreed damages clause requires that B pay to A a fixed sum of money upon B's breach of his contractual duty owed to A. As such, the court's review jurisdiction under the Penalty Rule conventionally arises where B must *pay money* upon *breach*. This chapter analyses those requirements.

After interpreting those requirements, the chapter canvasses recent case law that has extended the jurisdictional scope to include detriments other than the payment of money. The chapter then explores the breach requirement in three stages. First, the chapter outlines the present law in England. It then traces the recent history of the law in Australia. Second, the chapter provides an evaluation. It assesses the law in each jurisdiction, and the principled arguments that have been propounded to date in this area. Finally, the chapter presents a principled approach to the breach requirement that is capable of explaining Anglo-Australian private law as it stands in a conceptually satisfactory fashion.

1 Two elements

On an orthodox view, a court's review jurisdiction under the Penalty Rule arises where an agreed damages clause stipulates that B must pay a sum of money to A upon B's breach.[1]

Two points must be noted about the payment requirement. On the one hand, this requirement is not cast in rights language. It is more accurate to say that the clause must result in B incurring *an obligation* to pay A upon B's breach of contract. On the other hand, the traditional focus on an obligation *to pay* is too narrow. In *Andrews v Australia and New Zealand Banking Group Ltd*, the Australian High Court defined a penalty clause as imposing upon B an 'additional detriment' which benefits A.[2] Therefore, at the jurisdiction stage, a court must identify the relevant 'detriment' which B suffers. An obligation to pay money is merely an example. Then, at the validity stage, a court must determine whether any part of the detriment is 'additional'.

Accordingly, properly understood, the jurisdictional test contains two requirements. The first is that B must suffer a detriment. The second is that B's breach of contract must trigger that detriment. Each element is examined below.

---

[1] eg Michael Furmston (gen ed), *The Law of Contract* (4th edn, Butterworths 2010) [8.105].

[2] [2012] HCA 30, (2012) 247 CLR 205 [10].

2 Detriment

Recent case law has established that an obligation to pay money is merely a species of the detriment genus. Two other types of detriment suffice to satisfy the jurisdictional test, as follows.

A Transfer of rights

The first type is an obligation to transfer a proprietary right. In English law, *Jobson v Johnson* provides the classic example.[3] As Chapter 3 noted, in that case the relevant clause of the share sale contract provided that, if the purchaser defaulted in paying an instalment of the purchase price, he was obliged to retransfer to the vendor the proprietary rights to shares for a stipulated sum. The English Court of Appeal held that there was no distinction between a penalty clause that required B to pay money, and one that obligated B to transfer a proprietary right, either for no consideration or at an undervalue. In Nicholls LJ's words, 'In principle … there can be no difference between an obligation to pay a stipulated sum of money arising on a default and an obligation to transfer specified property arising on a default. The essential vice is the same in each case.'[4] In *Makdessi v Cavendish Square Holdings BV*, the English Court of Appeal endorsed this reasoning, holding that the Penalty Rule applies to a clause that 'requires a transfer of property … either for nothing or at an undervalue.'[5]

Australian law is similar. Two examples will be given. First, in *Ringrow Pty Ltd v BP Australia Pty Ltd*,[6] the relevant clause of an option deed provided that, if BP terminated a related contract for Ringrow's breach, BP was entitled to re-acquire from Ringrow the proprietary right to particular land for a stated purchase price. The High Court held that the Penalty Rule's jurisdiction extends to clauses stipulating that 'money's worth (including property) is transferable upon a particular event …'.[7] Second, in *Andrews*, the Australian High Court held that the relevant detriment is not restricted to an obligation to pay money, and that the English Court of Appeal's analysis is *Jobson* was exemplary. Distinguishing between an

---

[3] [1989] 1 WLR 1026 (CA).

[4] *Jobson v Johnson* [1989] 1 WLR 1026 (CA) 1042.

[5] [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [46].

[6] [2005] HCA 71, (2005) 224 CLR 656.

[7] *Ringrow Pty Ltd v BP Australia Pty Ltd* [2005] HCA 71, (2005) 224 CLR 656 [21].

obligation to pay and one to transfer a proprietary right would be unprincipled, as it would 'elevate form over substance'.[8]

If an obligation to transfer a *proprietary* right suffices, then an obligation to transfer a *personal* right must also do. Here, one cannot rationally distinguish between these two forms of rights. Further, as commercial parties usually intend that B will satisfy his obligation to pay money to A by an electronic funds transfer, and not in cash, arguably most agreed damages clauses require the transfer of a personal right.

Accordingly, a relevant detriment exists if the agreed damages clause imposes upon B an obligation to pay a sum or to transfer a personal or proprietary right to A. This will satisfy the first limb of the jurisdictional test.

B Extinguishment of rights

I Preliminaries

The second type of detriment involves the extinguishment of rights. Three preliminary points arise.

To start, an extinguishment occurs when a party gives up a pre-existing right that she held before the story began. For example, a party may hold a right to be paid money. That party may then enter a contract containing a clause whereby she 'contracts out' of her pre-existing right to be paid. Consequently, once the event triggering the clause occurs, that party's pre-existing right is destroyed for the future. Her entitlement is extinguished.

Next, a counterfactual test applies to determine whether an extinguishment has happened. One must ask whether, but for the relevant clause, a party would have held a right against another party (for instance, a right to be paid money). If the answer is 'yes', then an extinguishment has occurred.

Lastly, when discussing the issue of extinguishing rights, some Anglo-Australian judges and commentators have used the language of 'forfeiture'. However, to increase precision and avoid confusion, in this context the forfeiture terminology is best avoided.

II The law

[8] *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [13].

Leading academics have argued that the Penalty Rule is exclusively concerned with executory promises, and so it does not apply to benefits 'liable to forfeiture on breach'.[9] Not so.[10] An analysis of the English case law, and then the Australian case law, follows.

The leading English case is *Gilbert-Ash (Northern) Ltd v Modern Engineering (Bristol) Ltd*.[11] The facts concerned a building contract between a main contractor and a sub-contractor. Clause 14 of the sub-contract stated that the main contractor was entitled to 'suspend or withhold [the] payment of any monies due' to the sub-contractor if the latter breached the sub-contract's terms.[12] The House of Lords held that the clause fell within the court's review jurisdiction. According to Lord Salmon, a court is entitled to review a clause which, upon B's breach, would enable A to suspend or withhold the payment of money due from A to B.[13]

The English Court of Appeal has consistently supported the decision. For example, in *Firma C-Trade SA v Newcastle Protection and Indemnity Association (The Fanti)*, the Court followed the decision, holding that there is 'no distinction between withholding or disentitling a person to a sum of money which is due to him and requiring him to pay a sum of money …'.[14] Further, in *Makdessi*, the Court cited the decision in support of the proposition that the Penalty Rule applies to a clause that 'disentitles the contract breaker to receive a sum which would otherwise be due to him'.[15]

Stevens and McFarlane have argued that the agreed damages clause in *Gilbert-Ash* 'attempted to remove the subcontractor's accrued right to be paid if he committed any breach of contract … It was, therefore, functionally equivalent to a penalty payable upon breach of contract'.[16] This passage makes two fundamental points.

The first is that a clause entitling A to suspend or withhold the payment of money due from A to B, or disentitling B to receive a payment otherwise due to B, is in fact a clause that

[9] JW Carter and others, 'Contractual Penalties: Resurrecting the Equitable Jurisdiction' (2013) 30 JCL 99, 100. See also *Carter on Contract* [43-165] (service 37, April 2014).

[10] Sarah Worthington, 'What is Left of Equity's Relief Against Forfeiture?' in Elise Bant and Matthew Harding (eds), *Exploring Private Law* (CUP 2010) 270 cf Edwin Peel, 'Unjustified Penalties or an Unjustified Rule Against Penalties?' (2014) 130 LQR 365, 366 ('room for argument' exists).

[11] [1974] AC 689 (HL).

[12] *Gilbert-Ash (Northern) Ltd v Modern Engineering (Bristol) Ltd* [1974] AC 689 (HL) 703.

[13] *Gilbert-Ash (Northern) Ltd v Modern Engineering (Bristol) Ltd* [1974] AC 689 (HL) 723.

[14] [1989] 1 Lloyd's Rep 239 (CA) 262, 265, revd [1991] 2 AC 1 (HL).

[15] [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [46].

[16] Robert Stevens and Ben McFarlane, 'In Defence of *Sumpter v Hedges*' (2002) 118 LQR 569, 581.

extinguishes B's accrued right to be paid by A.[17] So, in *Gilbert-Ash*, clause 14 satisfied the definition of an extinguishment, as it effectively provided that, upon the sub-contractor's breach, it would give up its pre-existing claim right against the main contractor to be paid for services rendered. Clause 14 also passed the counterfactual test as, but for the clause, the sub-contractor would have held that claim right against the main contractor.

The second point relates to functional equivalency. Clearly, a clause requiring B to transfer a proprietary right to A for no consideration is functionally equivalent to one providing for B to pay money to A. Why? Because the former clause operates to leave B factually worse off, and to leave A factually better off, to the same extent as the latter clause. The same reasoning applies to the clause in *Gilbert-Ash*. Thus, clause 14 attempted to cancel the subcontractor's accrued right to be paid. Accordingly, clause 14 was functionally equivalent to an agreed damages clause requiring payment as the clause operated to leave the sub-contractor factually worse off, and to leave the main contractor factually better off, to the same extent as a clause requiring the sub-contractor to pay money to the main contractor. Put differently, a clause entitling A to extinguish an accrued right that B holds against A is functionally equivalent to a clause providing for B to pay money to A.

In Australia, it has been held that a provision that purports to extinguish B's right to receive moneys earned before termination falls within the review jurisdiction.

The authority in point is *Bysouth v Shire of Blackburn and Mitcham (No 2)*.[18] A contractor and Shire entered into a road-making contract. The relevant clause stipulated that if the contractor breached, the Shire obtained the right to terminate the contract and to extinguish the contractor's right to be paid for work he had completed. The contractor breached. The Shire purported to terminate the contract, and exercise its right to extinguish. In response, the contractor contended that the clause was penal. The Full Court of the Victorian Supreme Court held that the clause was reviewable. Irvine CJ held that the clause involved 'a forfeiture of existing rights'.[19] However, this only related to moneys due for work that had been certified. His Honour concluded that the 'forfeiture of the moneys already earned'[20] fell within the court's review power. In rights language, then, the Penalty Rule applies where upon breach a clause seeks to extinguish B's *accrued* contractual right to receive a payment from A.

[17] See *Lancore Services Ltd v Barclays Bank plc* [2009] EWCA Civ 752, [2010] 1 All ER 763 [29]; *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [120].

[18] [1928] VLR 562 (VSCFC).

[19] *Bysouth v Shire of Blackburn and Mitcham (No 2)* [1928] VLR 562 (VSCFC) 573.

[20] *Bysouth v Shire of Blackburn and Mitcham (No 2)* [1928] VLR 562 (VSCFC) 574. See also 585.

Later, in *Interstar Wholesale Finance Pty Ltd v Integral Home Loans Pty Ltd,* the New South Wales Court of Appeal assumed that the 'forfeiture of fully earned' contractual rights engaged the Penalty Rule.[21] Notably, in *Andrews*, the High Court seemingly endorsed the Court of Appeal's view that an extinguishment of an accrued contractual right suffices to engage the rule.[22]

This analysis clears up a misunderstanding. Some courts have previously held that the category of *extinguishment* of rights is conceptually distinct to that of *withholding* rights. For example, the Australian High Court has held that the Penalty Rule is engaged by an extinguishment *or* where a party 'suffer[s] the retention of money or property'.[23] Unfortunately, a lack of attention to rights creates double vision: the latter is merely an instance of the former. A clause that purports to extinguish an accrued right is a putative 'penal forfeiture clause', and falls within the jurisdiction.[24]

III A proprietary dimension

If an obligation to extinguish an accrued contractual right to be *paid* engages the Penalty Rule, will an obligation to give up an accrued contractual right to be transferred *property* also suffice? It must. Consider the following hypothetical clauses.

Clause 1 states that, if B breaches his obligation to do act X by Wednesday, B is obliged to pay to A the sum of £500. These facts represent the 'core case'. It is clear that B's obligation to *pay* money, or B's obligation to *transfer* a *personal* right, to A constitutes a relevant detriment.

Clause 2 stipulates that, if B breaches his obligation to do act X by Wednesday, B is obliged to transfer to A the proprietary right to shares worth £500. Given the Court of Appeal's decision in *Jobson* and the High Court's judgment in *Andrews*, this fact pattern falls within the Penalty Rule jurisdiction. In principle, it does so because an obligation to transfer a proprietary right is substantively identical to an obligation to transfer a personal right. Accordingly, B's obligation to *transfer* a *proprietary* right to A constitutes a sufficient detriment.

Clause 3 provides that, if B breaches his obligation to do act X by Wednesday, A is entitled to withhold payment to B of £500, a payment to which B has an accrued right. By

---

21 [2008] NSWCA 310, (2008) 257 ALR 292 [93], [104].

22 *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [30].

23 *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [49].

24 *Liberty Life Assurance Co Ltd v Sheikh* The Times, 25 June 1985 (CA).

reason of the House of Lords's decision in *Gilbert-Ash* and the Full Court's decision in *Bysouth*, this factual scenario engages the rule. In principle, clause 3 is functionally equivalent to clause 1: both clauses will operate to leave B factually worse off to the tune of £500, whilst leaving A factually better off to the value of £500. Accordingly, the *extinguishment* of B's *personal* right to payment—or B's right to have transferred money—constitutes a relevant detriment.

Clause 4 states that, if B breaches his obligation to do act X by Wednesday, A is entitled to withhold transferring to B the proprietary right to shares worth £500, a transfer to which B has an accrued right. In other words, A is entitled to extinguish B's accrued right to be transferred the shares.[25]

Examining first the case law, in *Interstar*, the New South Wales Court of Appeal held that it was 'a small step from accepting that the doctrine applies to a transfer of property … to applying it to [a] forfeiture of property …', and as such the court assumed that the Penalty Rule applies to the 'forfeiture of rights or property'.[26] This passage is difficult. It is best interpreted as follows. An obligation to transfer a proprietary right (under clause 2) is analogous to an obligation to extinguish a right to have transferred a proprietary right (under clause 4). Hence, both situations properly fall within the review jurisdiction.

Moving to the principled arguments, in the context of obligations to transfer, Anglo-Australian private law does not distinguish between personal and proprietary rights. Rationally, it must follow that the law does not refuse to review an obligation to extinguish merely on the basis that it connected to a proprietary right. In other words, when applying the jurisdictional test, the distinction between personal and proprietary rights is not a meaningful one.

It follows that, properly understood, the *extinguishment* of B's right to have transferred a *proprietary* right—or B's right to have transferred a benefit in kind—constitutes a sufficient detriment.

Consequently, a relevant detriment exists if: B holds a right against A that obligates A to transfer money or property to B; and the agreed damages clause provides A with a right to extinguish B's right to the payment or to have transferred the property, and imposes upon B

---

25 In some cases, if B holds an *accrued* contractual right to the transfer of shares, then B in fact may hold a proprietary right to the shares (as the beneficiary under a trust). If so, then pursuant to clause 4, A would be entitled to extinguish B's *proprietary* right to the shares. This possibility need not detain us further.

26 *Interstar Wholesale Finance Pty Ltd v Integral Home Loans Pty Ltd* [2008] NSWCA 310, (2008) 257 ALR 292 [104].

the concomitant obligation to give up the particular right. This will satisfy the first limb of the jurisdictional test.

3 Breach

It is universally agreed that the court's power to review an agreed damages clause arises upon B breaching a contractual duty owed to A. It has long been debated whether a rights infringement also delimits the court's review jurisdiction. Until recently, Anglo-Australian private law appeared united on this issue. However, the High Court's decision in *Andrews* has seemingly changed matters. The extent of the change effected is a difficult question.

A Two views

I England

In McGregor's view, it is a 'self-evident' proposition that an English court's jurisdiction can only arise if the event triggering the relevant detriment is B's breach of contract.[27] Thus, the 'breach requirement' doubles as a 'breach limitation'. Judges and jurists frequently cite three cases as establishing the breach limitation: *Export Credits Guarantee Department v Universal Oil Products Co*,[28] *Jervis v Harris*[29] and *Alder v Moore*.[30] Each is considered in turn.

Chapter 3 discussed *Export Credits*. The (simplified) facts were these. Procon contracted to construct a refinery for Provincial. Export Credits financed the project, acting as guarantor. Procon agreed with Export Credits that, if the guarantee was called upon at a time when Procon was in breach of its contract with Provincial, Procon would pay Export Credits a sum equivalent to that paid out under the guarantee. Export Credits paid out £39 million under the guarantee. At the time, Procon was in breach of its contract with Provincial. Export Credits demanded that Procon reimburse it. Procon claimed the relevant clause was penal.

Agreeing with the courts below, the House of Lords held the clause was unreviewable. Lord Roskill reasoned that the rule did not apply because the clause 'provided for payment of money on the happening of a specified event other than a breach of contractual duty owed by

---

[27] Harvey McGregor, *McGregor on Damages* (19th edn, Sweet and Maxwell 2014) [15-009].

[28] [1983] 1 WLR 399 (HL).

[29] [1996] Ch 195 (CA).

[30] [1961] 2 QB 57 (CA).

the contemplated payer to the contemplated payee'.[31] This statement provides the judgment's 'central' proposition.[32]

In *Jervis*, the facts involved a lease. Clause 2(7) provided that the tenant was required to maintain the premises in good condition. Clause 2(10) stated that if the tenant failed to remedy any defects or want of repair after receiving a written request to do so, the landlord would be entitled to enter the premises; to carry out the rectification or repair works; and to recover the cost of the work from the tenant. The landlord requested in writing that the tenant repair the premises. The tenant failed to comply. The landlord intended to carry out the repairs, but the tenant refused to allow access. The landlord sought to enforce his right under clause 2(10). The tenant argued that clause 2(10) was penal.

The English Court of Appeal unanimously held that the tenant could not invoke Court's review jurisdiction. Millett LJ held that it was 'well settled' that the event triggering the obligation to pay 'must be a breach of some other contractual obligation owed by [one party to the other]'.[33] On the facts, this was not the case. His Lordship held that the tenant's duty to pay for the repair costs was triggered by the landlord's expenditure on the repairs, not by the tenant's breach.[34] The breach, then, was simply part of the history.

In *Alder*, Moore was a professional footballer. He was injured, and certified as permanently disabled and unable to play football again. Consequently, Moore's insurers paid him £500 under his insurance policy. Pursuant to the policy terms, Moore gave the following declaration: 'I hereby declare and agree that I will take no part as a playing member of any form of professional football and that in the event of infringement of this condition I will be subject to a penalty of [£500]'.[35] Later, Moore began playing football again. The insurers sued for £500. Moore argued that the clause was penal.

A majority of the English Court of Appeal[36] held that the declaration imposed no 'obligation upon [Moore] never again to play professional football', so he committed no breach by resuming his football career.[37] The case did not involve a first promise not to resume play, with a second promise to pay £500 if the first promise was breached. Rather, Moore's only

[31] *Export Credits Guarantee Department v Universal Oil Products Co* [1983] 1 WLR 399 (HL) 402.

[32] *M and J Polymers Ltd v Imerys Minerals Ltd* [2008] EWHC 344 (Com Ct), [2008] 1 Lloyd's Rep 541 [41(i)].

[33] *Jervis v Harris* [1996] Ch 195 (CA) 206. See also 202-203.

[34] *Jervis v Harris* [1996] Ch 195 (CA) 204.

[35] *Alder v Moore* [1961] 2 QB 57 (CA) 62.

[36] Sellers LJ and Slade J, Devlin LJ dissenting.

[37] *Alder v Moore* [1961] 2 QB 57 (CA) 76. See also 65.

promise was to pay £500 *if he played again*. Therefore, the event that triggered the obligation to pay £500 was not a breach of contract. The clause was thus unreviewable. Devlin LJ dissented on the grounds that, in his view, the sum was payable upon a breach.

II Australia

The story of the breach limitation in Australia has unfolded in three stages: uncertainty, certainty, and certainty laced with uncertainty. Each stage is considered in sequence.

(1) Uncertainty

For many years Australia law was unclear as to two interrelated points. The first was the status of the *Export Credits* decision. The second was whether a breach limitation existed. Two Australian High Court cases will illustrate the point.

In *O'Dea v Allstates Leasing System (WA) Pty Ltd*,[38] Gibbs CJ and Brennan J each touched on the breach limitation. Gibbs CJ's judgment appears confused.[39] On the one hand, his Honour seemed to support a breach restriction, by citing *Alder* with apparent approval. On the other hand, the Chief Justice then appeared to question it, by leaving open the issue of whether the Penalty Rule applies when a hirer terminates a contract by exercising a contractual option. Conversely, Brennan J's view was pellucid: 'The balance of opinion in this Court has favoured the view that no question of penalty arises unless the obligation to pay arises upon [a] breach of contract.'[40]

In *AMEV-UDC Finance Ltd v Austin*,[41] five judges commented on the breach requirement. The two dissenting judges were split on the point. Dawson J confirmed that *Export Credits* was good law in Australia, and expressly accepted the breach limit.[42] By contrast, Deane J held that, properly understood, *Export Credits* did not support a breach limit; and after conducting an historical analysis his Honour rejected any breach restriction as being contrary to historical fact and general principle. Deane J suggested that the Penalty Rule jurisdiction extended to situations where a contractual liability exists:

---

38 (1983) 152 CLR 359 (HCA).

39 *O'Dea v Allstates Leasing System (WA) Pty Ltd* (1983) 152 CLR 359 (HCA) 367-68.

40 *O'Dea v Allstates Leasing System (WA) Pty Ltd* (1983) 152 CLR 359 (HCA) 390.

41 (1986) 162 CLR 170 (HCA).

42 *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 211.

> to pay or forfeit an amount or amounts either on or in default of the occurrence of an event which can be seen, as a matter of substance, to have been treated by the parties as lying within the area of obligation of the party liable to make the payment in the sense that it is his or her responsibility to ensure that the specified event does or does not occur …[43]

The majority judges' views were more opaque. Gibbs CJ held that the correctness of *Export Credits*, and the breach limit, was not in issue.[44] Mason and Wilson JJ's judgment is difficult. Their Honours referred without disapproval to *Export Credits*, but did not expressly endorse it. They then merely noted a difficulty in applying any breach limit, without expressing any view on whether the limit was correct as a matter of doctrine, or principle.[45]

Thus, at this stage it was unclear whether Australian courts supported the breach limitation. The law was unsettled.

(2) Certainty

The decision in *Interstar Wholesale Finance Pty Ltd v Integral Home Loans Pty Ltd* brought certainty to Australian law by confirming that the Penalty Rule only applied to clauses activated by a violation of contractual duty.[46]

At first instance, Brereton J concluded that the Penalty Rule did not apply only in cases of breach. His Honour found Deane J's reasoning in *AMEV* compelling. Accordingly, the judge adopted Deane J's jurisdictional test, holding that the Penalty Rule is engaged if B suffers a relevant detriment because an event occurs, or fails to occur, where substantively the parties have treated that event as lying within B's field of responsibility, in that B was responsible to see that the specified event did or did not occur.[47] In this thesis, the 'Responsibility Test' refers to this test.

On appeal, the New South Wales Court of Appeal reversed the decision. The Court's reasoning developed in five steps. First, the Penalty Rule was a rule of the common law, not equity. Second, the first instance court gave insufficient weight to *Export Credits*, and how clearly the House of Lords had expressed the relevant principle. Third, according to the Court of Appeal, Australian High Court authority supported a breach limit. Fourth, the trial judge

[43] *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 199.

[44] *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 174.

[45] *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 184.

[46] [2008] NSWCA 310, (2008) 257 ALR 292.

[47] *Integral Home Loans Pty Ltd v Interstar Wholesale Finance Pty Ltd* [2007] NSWSC 406, (2007) 2 BFRA 23 [71], [74], revd [2008] NSWCA 310, (2008) 257 ALR 292.

had failed to appreciate that Australian intermediate appellate court decisions had applied the Penalty Rule 'on the basis that payment being conditioned on a breach of contract [was] an essential element'.[48] Fifth, as a result, a first instance judge was not entitled to refashion the jurisdictional rule by removing the breach limitation. Only the High Court could enunciate a wider doctrine. Finally, it followed that in Australia the Penalty Rule contained a breach limit.[49] Emphasising *Export Credits*, the Court held that the doctrine was limited to clauses containing a breach trigger.

Immediately following *Interstar*, Australian appellate courts consistently held that the Penalty Rule did not apply where the trigger for the relevant obligation was an event other than a breach.[50] At this stage, the jurisdictional rule was clear.

(3) Certainty laced with uncertainty

The *Interstar* decision had brought certainty to an Australian law that was long unsettled. It was short-lived. The High Court's decision in *Andrews* reinstated the unease.

In *Andrews*, the applicants brought representative proceedings against an Australian bank. Under various clauses, the respondent bank charged its customers 'exception fees' for a variety of banking services it allegedly provided to them. The pertinent fees comprised 'honour', 'dishonour' and 'non-payment' fees customers incurred regarding deposit accounts; and 'over-limit' and 'late payment' fees incurred concerning credit card accounts. Generally speaking, the respondent bank levied the fees if the customer overdrew his or her account, or attempted to overdraw his or her account but the bank declined to process the transaction. The applicants claimed that the relevant clauses were penal. They sought repayment of the fees charged as 'moneys had and received' by the respondent bank.

At first instance, the Court held that the Penalty Rule jurisdiction applied only to the clauses that exacted exception fees upon a customer's breach of contract (namely, the late payment fees).

On appeal to the High Court, the only issue was whether the 'exception fee' clauses were capable of being reviewed as putative penalty clauses. Importantly, as regards some of these clauses, the bank did not charge the fees upon a breach, and the customer had no

---

[48] *Interstar Wholesale Finance Pty Ltd v Integral Home Loans Pty Ltd* [2008] NSWCA 310, (2008) 257 ALR 292 [126].

[49] *Interstar Wholesale Finance Pty Ltd v Integral Home Loans Pty Ltd* [2008] NSWCA 310, (2008) 257 ALR 292 [106].

[50] eg *Kowalczuk v Accom Finance* [2008] NSWCA 343, (2008) 77 NSWLR 205 [162]; *Ange v First East Auction Holdings Pty Ltd* [2011] VSCA 335, (2011) 284 ALR 638 [92]-[94].

'obligation or responsibility'[51] to avoid the occurrence of an event upon which the bank charged the fees. The Court unanimously held that, in Australia, a breach restriction does not exist; hence, a clause is not unreviewable merely because it does not contain a breach trigger, or is not connected to a breach.[52] Notably, the Court did not decide any issues of fact. That is, it did not determine whether the relevant provisions actually fell within the courts' review power, or whether they fell foul of any tests of validity. The High Court remitted those matters to the trial judge to decide.[53]

The High Court's analysis was entirely historical. In summary, it contained four steps.

First, the High Court established that before 1873, equity relieved against stipulations exacting a penalty upon an event occurring, or not occurring. That event, however, did not necessarily involve a breach of contract. The Court explained the point by reference to the rule's origins, and early development, as follows.[54]

As to the former, the Penalty Rule originated in equity, where the Court of Chancery intervened to relieve a party against the other enforcing a penal bond. In a penal bond, the obligor's obligation to pay depended upon a 'condition'. If the condition was fulfilled, the obligation was defeasible. However, equity courts did not use the term 'condition' to mean a (promissory) obligation. Rather, the term merely referred to the occurrence, or non-occurrence, of a particular event. That event did not necessarily involve the obligor breaching an obligation.[55]

As to the latter, the Court of Chancery later began to relieve against penalty clauses in simple contracts. However, the scope of the court's review jurisdiction remained the same.[56] In short, the High Court held that, historically, equity's jurisdiction to relieve against penalties was never confined to detriments incurred upon a breach of contract.

Second, the Court held that prior to 1873, the common law had never supplanted equity's jurisdiction. Once the Chancery Court began frequently granting relief against penalty clauses, the common law courts expanded the action in assumpsit, and began to give the same

---

[51] *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [22], [78].

[52] *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [45]-[46], [78].

[53] Upon being remitted back to the Federal Court, Gordon J held that only the clauses charging late payment fees were reviewable, and penal: *Paciocco v Australia and New Zealand Banking Group Ltd* [2014] FCA 35, (2014) 309 ALR 249 [4]-[5], [373]-[374].

[54] See also text to nn 11-15 in ch 2.

[55] *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [42]. See also [39].

[56] *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [44].

relief as the equity courts. However, the High Court stressed that this development of the legal Penalty Rule was merely 'procedural'; hence, it did not affect the equitable version of the rule.[57]

Third, in 1873, English parliament passed the Supreme Court of Judicature Act 1873. The Judicature system unified court procedure, not substantive doctrine. It followed that two separate Penalty Rules—one at law, the other in equity—remained. Consequently, in *Interstar*, the New South Wales Court of Appeal had erred, in that it had no basis to hold that the Penalty Rule 'is a rule of law not of equity'.[58]

Finally, given that the equitable doctrine prevails, the modern Penalty Rule is an equitable rule. As stated, the Court held that its jurisdictional test does not contain a breach limitation.

The High Court held that a violation of a contractual duty was not *necessary* to engage the rule. It did not clearly define the nature of the event that is *sufficient* to invoke the rule. However, the Court did enunciate three important principles.

To start, it held that the court's review power arises in the following circumstances: where, assessed substantively, a primary stipulation provides A with a benefit; a collateral stipulation states that B will assume a detriment to A's benefit if the primary stipulation fails; and the collateral stipulation provides a security for B satisfying the primary stipulation.

Next, the Court held that the Penalty Rule is not engaged if the 'prejudice or damage' A suffers from the primary stipulation failing is 'insusceptible of evaluation and assessment in money terms'.[59]

Lastly, the High Court held that the jurisdiction is not invoked if B incurs a relevant obligation as a result of an 'alternative stipulation' which provides for an 'additional obligation'. An 'alternative stipulation' is a clause stating that B must pay to A an additional sum, or the like, for a further right or benefit that A has agreed to provide to B.[60] For example, A may agree to provide to B an extra service for a greater fee. This type of clause, then, is in the nature of a legitimate option.

---

[57] *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [51], [53], [55], [61].

[58] *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [63].

[59] *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [11].

[60] *Paciocco v Australia and New Zealand Banking Group Ltd* [2014] FCA 35, (2014) 309 ALR 249 [5], [374].

The Court presented *Metro-Goldwyn-Mayer Ltd v Greenham*[61] as an example case. A distributor and exhibitor entered into a contract for the hiring of films for public showing. The contract provided the exhibitor with the right to screen a film once only. The agreement further stipulated that if the exhibitor showed the film without the distributor's consent, he would pay the distributor for each subsequent, unauthorised screening a sum equivalent to four times the standard hiring fee. Several unauthorised showings occurred. The distributor sued for the sums payable as a result of the additional screenings. The exhibitor argued that the clause was penal. The majority of the New South Wales Court of Appeal held the clause was unreviewable.[62] The clause did not provide for an exaction on breach, but rather granted an option to rescreen the film for a significantly increased price. In Jacobs JA's words, with which the High Court agreed:

> There is no right in the exhibitor to use the film otherwise than on an authorised occasion. If he does so then he must be taken to have exercised an option so to do under the agreement ... The agreement provides that he may exercise such an option in one event only, namely, that he pay a hiring fee of four times the usual hiring fee.[63]

In conclusion, the jurisdictional scope of the Penalty Rule in Australia is certain, and uncertain. It is certain in that a rights infringement does not delimit its jurisdictional scope. It is uncertain in that it is unclear as to what, if anything, restrains the court's review power.

B Evaluation

I English law

Peel has written that the breach limitation appears to have become 'firmly entrenched as a matter of English law'.[64] The limitation is indeed routinely accepted. However, a close examination of the three English cases apparently supporting the breach limitation is telling. It reveals that appearances can deceive, and that what is routinely accepted is challengeable. These cases are examined in turn.

---

61 [1966] 2 NSWR 717 (NSWCA).

62 *Metro-Goldwyn-Mayer Ltd v Greenham* [1966] 2 NSWR 717 (NSWCA) 723, 726-27.

63 *Metro-Goldwyn-Mayer Ltd v Greenham* [1966] 2 NSWR 717 (NSWCA) 723.

64 Edwin Peel, 'The Rule Against Penalties' (2013) 129 LQR 152, 153.

The first case is *Export Credits*. Most assume the decision is authoritative in England. However, the judgment's support for a breach restriction is diluted given that the outcome can be rationalised on two other bases.

On the one hand, it is arguable that the relevant clause was not an agreed damages clause at all; rather, it was an indemnity clause.[65] Thus, in *Export Credits*, the House of Lords adopted the English Court of Appeal's reasoning. The latter had held that the sum Export Credits claimed was 'a debt due under an indemnity', which created an 'entirely different' type of claim to that brought under an agreed damages clause.[66] Why? Because the two types of clauses are distinct in three respects: measure, quantifiability and timing.

As to measure, indemnity clauses involve a duty to pay compensation to redress *incurred* loss suffered,[67] whereas agreed compensatory damages involve a duty to pay a sum representing *anticipated* loss. An element of approximation exists only in the latter case. As a result, a claimant suing on a liquidated compensatory damages clause may recover more, or less, than her incurred loss. A claimant suing on an indemnity clause cannot.

As to quantifiability, the amount payable under an indemnity clause will *never* be quantifiable at the contract date, as it is wholly dependent on the claimant's incurred loss. Conversely, the amount payable as agreed compensatory damages is *necessarily* quantifiable at the contract date, this being a central feature of agreed compensatory damages clauses.[68] On a related note, given the lack of quantification, a claimant enforcing an indemnity clause must prove her incurred loss in the usual way. A claimant enforcing a liquidated compensatory damages clause needs not, this being a main advantage of these types of clauses.[69]

As to timing, a claimant's cause of action under an indemnity clause does not accrue until either she has incurred a loss,[70] or the sum she is due to pay is ascertained.[71] Conversely, a claimant may enforce a liquidated damages clause upon the triggering event occurring.

---

[65] *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 198-99 (Deane J); *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [47].

[66] *Export Credits Guarantee Department v Universal Oil Products Co* [1983] 1 Lloyd's Rep 448 (CA) 456, 461, affd [1983] 1 WLR 399 (HL).

[67] cf Rafal Zakrzewski, 'The Nature of a Claim on an Indemnity' (2006) 22 JCL 54, 55-56, 67, 70.

[68] Text to n 13 in ch 1.

[69] cf JW Carter and Wayne Courtney, 'Indemnities against Breach of Contract as Agreed Damages Clauses' [2012] JBL 555, 566.

[70] *Callaghan v Dominion Insurance Co Ltd* [1997] 2 Lloyd's Rep 541 (QB) 546.

[71] *County and District Properties Ltd v C Jenner and Son Ltd* [1976] 2 Lloyd's Rep 728 (QB) 734.

In summary, agreed damages clauses and indemnity provisions are conceptually distinct. An indemnity clause clearly is not subject to the Penalty Rule.[72] The facts of *Export Credits* involved precisely that type of clause. It follows, then, that the judgment was silent on the rule's jurisdictional scope, for properly it was not in issue.

On the other hand, assume for argument's sake that the relevant term *was* an agreed damages clause. If so, the 'central' proposition that encapsulated the breach restriction provided only one plank in the House of Lords's reasoning. Independently, Lord Roskill stressed that Export Credits was only seeking reimbursement for its expected loss. Indeed, his Lordship's judgment emphasised that the sum payable corresponded with the anticipated legally-recoverable loss—and hence it passed the tests of validity—perhaps more so than the breach point.[73] The decision was thus justifiable on that basis alone, divorced from any breach limitation reasoning.

The next case is *Jervis*. The decision is vulnerable, for at least three reasons.

First, if indemnity clauses and agreed damages clauses are truly distinct, then clause 2(10) was of the former type. It imposed a duty to pay compensation to redress incurred loss, and lacked the necessary quantification feature.[74] Accordingly, this undermines Millett LJ's observations regarding the rule's jurisdictional admit.

Second, the tenant's duty to pay for the repair costs under clause 2(10) could arise *only* if the tenant breached his obligation to keep the property in repair under clause 2(7). For the landlord to invoke his reimbursement right under clause 2(10), two conditions were necessary: the tenant must have breached his obligation under clause 2(7), and then the landlord must have exercised his right to enter the property and effect the necessary repairs under clause 2(10). Millett LJ acknowledged this. For example, his Lordship held that although the tenant's duty to pay was triggered by the landlord's expenditure on the repairs, for this event to occur there was must have existed an 'anterior failure of the tenant to repair'.[75] Accordingly,

---

72 *Total Transport Corp v Arcadia Petroleum Ltd (The Eurus)* [1996] 2 Lloyd's Rep 408 (Com Ct) 422, affd [1998] 1 Lloyd's Rep 351 (CA) cf *Yes Home Loans Pty Ltd v Afig Wholesale Pty Ltd* [2008] NSWSC 1017 [136]-[137].

73 cf *Export Credits Guarantee Department v Universal Oil Products Co* [1983] 1 WLR 399 (HL) 402 (breach) and 403 (correspondence). See *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [47].

74 *Jervis v Harris* [1996] Ch 195 (CA) 202 (redress incurred loss), 203 (lack of quantification).

75 *Jervis v Harris* [1996] Ch 195 (CA) 206. See also 204.

the two clauses were interdependent. That dependency should have sufficed to bring the clause 2(10) within the Court's purview.[76]

Finally, assume the Court of Appeal decided that it *did* have jurisdiction to review the clause. If so, then as in *Export Credits*, the result would have remained unchanged. On the facts, judged as at the contract date, the tenant's anticipated payment corresponded with the landlord's likely loss. Hence, the clause passed the tests of validity.

The final case is *Alder*. Three points arise.

To start, arguably the relevant clause was not an agreed damages clause or an option clause; rather, it was a consensual restitution clause. The insurer paid the £500 on the basis that Moore was permanently disabled. The clause was concerned to provide the insurer with a restitutionary right if that state of affairs was falsified.[77] So, the contractual right sought to replicate a claim right in unjust enrichment for a failure of basis. The decision is best explained in the following way: the insurer paid the £500 on the shared basis that Moore was permanently disabled; that basis failed; and the insurer recovered based on its contractual restitutionary right. It follows that the judgment is far from authoritative as regards the Penalty Rule's review jurisdiction. Properly understood, the question whether the rule applied did not arise.

Next, the Court of Appeal's decision that no obligation existed on the facts was suspect. The declaration contained the following words: 'I hereby declare and agree that I will take no part as a playing member … and that in the event of infringement of this condition I will be subject to a penalty of [£500]'. In dissent, Devlin LJ argued that the declaration's meaning was clear—the language was 'as plain as a pikestaff'[78]—in that it contained a promise. Lord Goff certainly thought so.[79] Indeed, if such words in an ordinary insurance contract do not create an obligation, it is difficult to see what words will do so.

Lastly, the judgment's support for a breach limitation is weakened given that the result can be justified on a different basis. Suppose the Court of Appeal had agreed to review the clause. In *Alder*, Sellers LJ held that the clause would have satisfied the tests of validity, as the £500 sum was a 'fair pre-estimate' of the insurer's anticipated loss.[80] Conversely, Devlin LJ

[76] *M and J Polymers Ltd v Imerys Minerals Ltd* [2008] EWHC 344 (Com Ct), [2008] 1 Lloyd's Rep 541 [41].

[77] *Alder v Moore* [1961] 2 QB 57 (CA) 65, 76.

[78] *Alder v Moore* [1961] 2 QB 57 (CA) 69.

[79] Robert Goff, 'Rewriting Contracts and the Court of Appeal' (1961) 24 MLR 637, 640.

[80] *Alder v Moore* [1961] 2 QB 57 (CA) 64.

effectively held that, as the insurer was always unlikely to suffer any loss from Moore's breach, the sum would fail any test of validity, and was therefore penal.[81] Leading commentators agree with the former analysis.[82] In fact, for two reasons, the latter analysis is correct.

The first relates to causation. The insurer was never going to lose the £500 *as a result of* Moore playing football again, because his breach of contract would never cause the insurer's payment according to a 'but for' test of causation. So, if Moore had committed no breach, the insurer would still have paid the sum, and he would have been entitled to keep the money.

The second reason concerns the point of comparison. Sellers LJ held that the £500 sum passed the tests of validity because his Lordship compared that amount with the position the insurer would have been in if the contract not been made.[83] However, Sellers LJ should have compared that amount with the position the insurer would have occupied if the contract had been performed.[84] Had the contract been performed, the insurer would have paid out £500 and recovered nothing. Accordingly, assessed at the contract date, recovering the £500 'penalty' would place the insurer into a far better position than if the contract had been performed. It followed that the clause was a penalty by name, and by its true nature.

However, the story does not end there. If the Court had declared the clause invalid, Moore would not have been obligated to make the £500 payment. However, per hypothesis, Moore would have breached the contract. Accordingly, the insurer would have been entitled to terminate the contract for breach and claim restitution of its original £500 payment based on a total failure of basis, given that Moore was not permanently disabled. On this analysis, then, the outcome would have been the same: a court order that Moore pay the insurer £500.

It follows that the leading English cases allegedly supporting the breach restriction are challengeable. Also, other authorities exist which militate against any such limit. For example, in *Bridge v Campbell Discount Co Ltd*, Lord Denning held that a breach restriction simply did not exist, and Lord Devlin doubted that a breach was necessary in all cases.[85]

There is one final group of cases suggesting that an English court may review a clause that does not provide for a detriment upon the violation of a contractual duty. The relevant clause is a 'disguised' agreed damages clause: that is, a clause drafted in an attempt to evade the

---

[81] *Alder v Moore* [1961] 2 QB 57 (CA) 72-74.

[82] eg Michael Furmston (gen ed), *The Law of Contract* (4th edn, Butterworths 2010) [8.111] fn 5.

[83] *Alder v Moore* [1961] 2 QB 57 (CA) 64, 66 cf 73.

[84] *Robinson v Harman* (1848) 1 Exch 850, 855; 154 ER 363, 365.

[85] [1962] AC 600 (HL) 631, 633-34.

Penalty Rule jurisdiction. If such a clause is invalid, it is known as a 'disguised' penalty clause. Three cases suggest that a 'disguised' agreed damages clause will not escape review.

First, in *Interfoto Picture Library Ltd v Stiletto Visual Programmes Ltd*,[86] Interfoto supplied photographic transparencies to Stiletto. It did so on conditions requiring Stiletto to return the transparencies within 14 days from the delivery date, otherwise a 'holding fee' of £5 per day would be charged for each transparency retained beyond that period. Stiletto returned the transparencies well after the deadline. Interfoto sought to recover the holding charge, which totalled approximately £3,780. The English Court of Appeal held that the clause was not incorporated into the contract, so the claim failed. In reaching this conclusion, Bingham LJ suggested that the clause was potentially 'challengeable as a disguised penalty clause'.[87] This was so despite the fact that Stiletto did not breach the contract by returning the transparencies late.

Second, in *Euro London Appointments Ltd v Claessens International Ltd*,[88] Chadwick LJ held that the term in *Interfoto* was vulnerable to attack under the Penalty Rule. His Lordship continued, 'In referring to the clause as "a disguised penalty clause", Bingham LJ was pointing out that, if … the true nature of the bargain [was that it involved a contractual obligation], the clause could not be taken outside the [Penalty Rule] by presenting the charge as a "holding fee".'[89]

Finally, in *Office of Fair Trading v Abbey National plc*, Lord Phillips suggested that a party could not evade review by ensuring that 'the contingencies that triggered liability to pay … did not constitute breaches of contract'.[90]

These dicta suggest that, in English law, the review jurisdiction may extend beyond clauses containing a breach to those that are purposively drafted to avoid one. For example, on this view, the rule will catch a clause that seeks to avoid the breach requirement by presenting B with an 'option' to perform in one way rather than another. An obligation masquerading as an option cannot escape.

II Australian law

---

[86] [1989] QB 433 (CA).

[87] *Interfoto Picture Library Ltd v Stiletto Visual Programmes Ltd* [1989] QB 433 (CA) 445-46.

[88] [2006] EWCA Civ 385, [2006] 2 Lloyd's Rep 436.

[89] *Euro London Appointments Ltd v Claessens International Ltd* [2006] EWCA Civ 385, [2006] 2 Lloyd's Rep 436 [28].

[90] [2009] UKSC 6, [2010] 1 AC 696 [83].

The High Court's judgment in *Andrews* shows a mastery of legal history. It is an impressive archaeological work. However, for five reasons, the reasoning is questionable.

First, the most important feature of the High Court's reasoning is what it is missing from it. Orthodoxy dictates that a breach limitation exists. This raises two questions. Does the restriction really exist? If it does exist, should the jurisdiction be extended beyond breach? To answer these questions, it is necessary to identify the Penalty Rule's normative justification. To explain, it is generally accepted that the breach restriction is a key feature of the review jurisdiction. However, one cannot know if that is truly so without a theoretically appropriate understanding of the rule. The first task must be to understand why the rule exists. It is only once we determine this that we can evaluate whether the rule's putative core features are defensible as such. As Furmston has argued, 'In a rational world, the formulation of the rule should be related to an enquiry into the rationale of the rule.'[91] Without identifying the rule's theoretical rationale, it is impossible to justify or impeach the breach limit, or offer proper guidance as to how to extend the jurisdictional rule beyond breach. These tasks cannot be performed in the abstract. Notably, the High Court's judgment does not contain a hint of this discussion. Accordingly, the court's reasoning failed to meet the standards required of any authoritative judgment.

Second, the Court's reasons in support of its conclusion were entirely historical. It argued that historically a rights infringement was not required to invoke the Penalty Rule, and so it follows that one is not required today. The premise may well be correct. The conclusion, however, does not follow from the premise. This argument commits a 'genetic fallacy'. As Beever has written, 'I was an infant, but I hope it can be agreed that infantilism is not an essential part of my nature.'[92] When the rule first arose, a breach may not have been a necessary element of the jurisdictional test. However, the rule has evolved. The question is whether a breach is a necessary element of the *current* doctrine. To answer that question, it is insufficient solely to rely on a state of affairs that existed several centuries ago. As the rule has evolved, so may have its essential elements. Legal history is very important. But the nature of a modern day doctrine is not shackled to it. Otherwise, the private law forever would be captured by its past.

Third, the High Court's purely historical reasoning is flawed, for it is incomplete. Thus, the judgment in *Andrews* closely traces the origins and development of the Penalty Rule

---

[91] MP Furmston, 'Contract Planning: Liquidated Damages, Deposits and the Foreseeability Rule' (1991) 4 JCL 1, 7.

[92] Allan Beever, *The Law of Private Nuisance* (Hart Publishing 2013) 40.

up to and including the Judicature reforms. It then pays scant attention to over 140 years of case law that Anglo-Australian courts have handed down since. A purely historical account should seek to explain how and why the rule developed the way it has to the present day; it should 'reveal the law's causal history'.[93] Stopping the analysis at an arbitrary mid-way point in the modern rule's development does not do this.

Where the judgment does engage with the modern leading cases, its analysis is at times artificial. For example, the Court argued that, in *AMEV*, Mason and Wilson JJ 'overlooked' the fact that the Judicature system effected only procedural change.[94] This is implausible. Mason and Wilson JJ's judgment referred on several occasions to the Judicature reforms, including the following: 'The advent of the Judicature system … hastened the demise of equity's separate jurisdiction to relieve against penalties.'[95] Further, their Honours would no doubt have read Deane J's dissenting judgment, which precisely discussed the issue of fusion and the Judicature system.[96] Accordingly, in *Andrews*, the High Court's reasoning was not analysing history, but reinventing it.

Fourth, the Court held that the review jurisdiction is not invoked if the loss A suffers from the primary stipulation failing cannot be evaluated or assessed in money terms. However, the fact A's anticipated loss is incapable of evaluation or assessment is properly not relevant to the jurisdiction issue. Rather, it is relevant to the *tests of validity*. In particular, under the Impossibility Test, the fact precise pre-estimation of loss is 'almost an impossibility' will likely lead to a court validating the clause. In fact, in *Andrews*, the High Court acknowledged this. In discussing penal bonds, it held that equity 'would give relief' where the loss was capable of assessment, but 'it would not' where compensation could not be assessed, and that this distinction 'developed into the modern distinction between penalties and liquidated damages'.[97]

Finally, the High Court relied upon two distinctions that are not intellectually satisfactory: between a 'primary' and a 'collateral' stipulation, and between a 'collateral' and an 'alternative' stipulation.

---

[93] Stephen A Smith, *Contract Theory* (OUP 2004) 4.

[94] *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [68].

[95] *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 191.

[96] *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 201.

[97] *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [41].

As to the former, to apply the distinction between primary and collateral stipulations a court will need to ascertain the agreement's characteristic stipulation. It will then label that stipulation as primary, and the other collateral to it. In many cases, it will be far from obvious which stipulation is the characteristic one. The Court offered no indicia for identifying a collateral obligation. The leading modern decisions—such as *AMEV* and *Ringrow*—do not even hint at the distinction, let alone provide guidance on how to identify a collateral stipulation. In these 'hard cases', then, the distinction is useless. In other cases, the distinction clearly will 'not readily accommodate the … facts', such that a court will be forced either to use 'untidy' artificial reasoning to shoehorn the facts into the distinction, or to jettison it altogether.[98]

As to the latter, the distinction between collateral and alternative stipulations is completely obscure. The High Court held that *Metro-Goldwyn-Mayer* illustrated the difference between the two types of stipulations. However, in that case, Jacobs JA held that one reason for the 'very large' increase in the 'price' may well have been to 'strongly discourage' any additional showings.[99] Accordingly, one might easily describe the promise to pay the inflated 'hiring fee' as being in the nature of a security for, and made to deter any deviation from, the commitment only to screen the film in accordance with the contractual regime. That, of course, precisely satisfies the High Court's definition of a collateral stipulation: namely, a stipulation 'in the nature of a security for and *in terrorem* of the satisfaction of the primary stipulation …'.[100]

In conclusion, the English precedents do not definitively decide that a breach restriction exists, but nor do they set out the precise limits of the jurisdictional rule. Conversely, the Australian case law has conclusively determined that breach does not constitute a limit, but it provides no guidance as to the review power's proper scope. It is thus necessary to examine the principled arguments in favour of expanding the jurisdiction to see if they support extending the rule, and if so to what extent.

III Principled arguments

Few support the breach restriction. One supporter is Carter, who describes the 'logic of the approach' as 'largely unassailable', from two perspectives.[101] First, it is said that a breach

[98] *GWC Property Group Pty Ltd v Higginson* [2014] QSC 264 [27], [37].

[99] *Metro-Goldwyn-Mayer Ltd v Greenham* [1966] 2 NSWR 717 (NSWCA) 723.

[100] *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [10].

[101] *Carter on Contract* [43-160] (service 37, April 2014).

restriction makes sense due to the nature of the provision, being one providing for agreed *damages*. Second, the writer asserts that a breach limitation is sound given the concept of a 'penalty', which is said to be 'an excessive imposition for not doing something, or for doing something not permitted'.[102]

More commonly, it is argued that the breach limitation causes injustice. Three principled arguments are said to support this proposition.

Lord Denning articulated the first argument in *Bridge*. His Lordship argued that the breach restriction leads to the 'absurd paradox'[103] that the courts will assist B by reviewing the clause if he breaches his contract with A, but not if he complies with its terms. In *AMEV*, Deane J agreed, holding that this 'illogical' paradox was to be 'denounced'.[104] For some scholars, it presents a 'major problem' that seemingly cannot be solved.[105] In a similar vein, appellate Anglo-Australian courts have recently labelled the breach limitation as 'anomalous'[106] and 'arbitrary'.[107]

It is worth pausing to ask why the breach limit presents a paradox, and why it is absurd. It turns out the limit may be neither, for at least two reasons.

For one, if the doctrine's proper normative justification mandates a breach requirement, then so be it. That will be the only way to render the Penalty Rule coherent. There is no paradox—let alone an absurd one—if the breach requirement is logically and rationally defensible by reference to the rule's correct normative explanation. Accordingly, without first identifying the rule's theoretical rationale, Lord Denning's allegation that a 'paradox' exists is empty.

For another, without a proper theoretical justification for the Penalty Rule, *every* jurisdictional rule can be labelled arbitrary, illogical, or absurd. Consider, for instance, the High Court's proposed distinction in *Andrews* between collateral and alternative stipulations. In the abstract, that distinction might be criticised as anomalous, as it seemingly means the rule will assist a party who makes two promises ('I promise that I will keep my promise'), but not one who only makes one promise ('I promise to perform by doing X or Y'). One can only

---

102 *Carter on Contract* [43-160] (service 37, April 2014).

103 *Bridge v Campbell Discount Co Ltd* [1962] AC 600 (HL) 626.

104 *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 198, 204-205.

105 Sirko Harder, 'The Relevance of Breach to the Applicability of the Rule Against Penalties' (2013) 30 JCL 52, 68-69.

106 *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [123].

107 *Bartercard Ltd v Myallhurst Pty Ltd* [2000] QCA 445 [2].

properly judge whether any jurisdictional test is paradoxical based on the rule's true justification.

The second argument is that the need to show a breach of duty leads to doctrinal distortion. Gullifer has argued that the Penalty Rule's jurisdictional inadequacy may lead courts to manipulate other doctrines to address the perceived injustice.[108] The writer gives the following example. Due to the breach requirement, a court may lack control over a disproportionate payment clause in a finance lease. This may cause it to grant relief against forfeiture of a possessory right, such that the payment is avoided but the agreement is kept on foot, even though termination is more appropriate and would follow, if the Penalty Rule applied. The court may do so *merely because* the clause is unreviewable under the rule.

This view is seriously flawed. On the one hand, the writer's argument assumes that courts will use doctrine X to make up for the inadequacies in doctrine Y. But the fact is that doctrine X will have certain elements, and a court will only grant relief if those elements are satisfied. If one truly believes that judges comply with the rule of law, and courts apply legal doctrines to reason *to* a conclusion, rather than reason *from* a conclusion, then this argument simply does not work. Each doctrine will be applied on its own terms, and relief granted or withheld accordingly. On the other hand, if courts do in fact (illegitimately) manipulate legal rules to reach a preordained conclusion, then the elements of any doctrine—including its 'inadequacies'—are irrelevant in any event: in every case, it is palm tree justice that decides private law disputes.

The final argument is this. It is said that the breach restriction permits a clever drafter to circumvent the review jurisdiction through various devices.[109] So, the drafter may adopt an avoidance technique whereby she uses a type of clause that *in form* does not involve breach to cater for what *in substance* amounts to an obligation triggered by a breach. Importantly, these types of clauses are not illegitimate *per se*. A party that chooses to adopt one of these types of clauses does not necessarily seek to evade the Penalty Rule jurisdiction. Rather, the form the clause takes allows it be put to that use, in the right hands.

A couple of examples might help.

---

[108] Louise Gullifer, 'Agreed Remedies' in Andrew Burrows and Edwin Peel (eds), *Commercial Remedies: Current Issues and Problems* (OUP 2003) 201.

[109] Carmine Conte, 'The Jurisdiction to Relieve against Penalties and Forfeitures: Time for a Rethink' (2010) 126 LQR 529, 531.

One structure that a party may use in an attempt to evade the jurisdiction is that of an option clause. On the orthodox view, the court's review power is not invoked if a party exercises an option to perform in one way rather than another, for no breach exists in any event.[110] As such, a party might use an option clause structure in an attempt to avoid review. For instance, assume Y hires out DVDs. The usual hire rate for new release films is £10 for one night. X hires such a film on Monday. A clause may state that X has the 'option' of either returning the DVD on Tuesday and paying £10, or returning it on Wednesday and paying £40.

Another structure that a party may use to try and avoid the review jurisdiction is that of an incentive payment clause.[111] These clauses come in various styles.

On the one hand, such a clause may offer a 'discount' for prompt payment. Any clause of this nature will not engage the rule if its jurisdictional test contains a breach restriction.[112] For example, suppose A loans money to B. The market interest rate for an ordinary commercial loan is 5 per cent. A clause may provide that B is to pay A an interest rate of 25 per cent on any outstanding sums, subject to a 'discount' of 20 per cent if B makes his repayments on or before the specified dates.

On the other hand, an incentive payment clause may provide for a 'bonus' for early completion. Peel gives the following example.[113] Assume that C agrees to construct a building for D. Compare then the following two clauses. The first clause states that C promises to complete the building by 1 January; D agrees to pay £1,000,000; and C agrees to pay £50,000 for each month it delays. The second clause imposes the following regime: if C completes the building on or before 1 January, D will pay £1,000,000; if C completes in the period of 2 January to 1 February, D will pay £950,000; if C completes in the period of 2 February to 1 March, D will pay £900,000; and so on. The Penalty Rule applies to the first provision. If a breach limitation exists, it does not apply to the second.

Allowing a party to avoid the rule frustrates the private law's stance against penalty clauses, emptying the doctrine of any utility. This argument has force. It will be explored further below.

---

[110] *Export Credits Guarantee Department v Universal Oil Products Co* [1983] 1 Lloyd's Rep 448 (CA) 456, affd [1983] 1 WLR 399 (HL); *Love v Brien* [2012] WASC 457 [65], affd [2013] WASCA 280.

[111] *The General Trading Co (Holdings) Ltd v Richmond Corp Ltd* [2008] EWHC 1479 (Com Ct), [2008] 2 Lloyd's Rep 475 [114].

[112] *Astley v Weldon* (1801) 2 Bos & P 346, 353; 126 ER 1318, 1323-24; *O'Dea v Allstates Leasing System (WA) Pty Ltd* (1983) 152 CLR 359 (HCA) 366-67.

[113] Edwin Peel, 'The Rule Against Penalties' (2013) 129 LQR 152, 154.

There is one more principled point to consider. It might be contended that a breach restriction is untenable given that, in some cases, a rights infringement strictly is not the event triggering the relevant detrimental obligation, and yet the clause is reviewable upon orthodox principles.

This situation in play is this. Suppose a clause states that A has the right to terminate the contract pursuant to an express contractual right arising upon B's non-repudiatory breach. A separate clause states that, if A does so, B must pay agreed damages to A. Strictly, when such an agreed damages is engaged, it is A's election to exercise her right to terminate, not B's breach, that causally triggers B's obligation to pay. If a breach limitation applies, it seemingly follows that the clause is unreviewable.

This is not the law. In Australia and in England, the clause falls within the court's review power.[114] This is because, in principle, Anglo-Australian courts take a substantive approach to the jurisdictional issue.[115] The reasoning here is that, in substance, it is the underlying breach that triggers B's duty to pay, such that the agreed damages clause is reviewable. The approach constitutes treating 'termination of an agreement upon breach in the same way as the breach itself for the purpose of determining whether a stipulated payment is capable of amounting to a penalty…'.[116] In short, termination based upon a breach is essentially tantamount to a breach, and engages the jurisdiction.

C A principled approach

I The approach

At present, Anglo-Australian private law on the breach limitation is unclear. It can confidently be said that, on its face, English law supports the restriction, and Australian law does not. Beyond that, the law is opaque. Some principled arguments have purchase, but they do not present a coherent, justifiable approach to this issue. This section presents a principled approach to the breach requirement. It is an approach that allows one to comprehend the present Anglo-Australian law in a conceptually satisfactory fashion. It proceeds in four stages.

---

114 *Cooden Engineering Co v Stanford* [1953] 1 QB 86 (CA) 96-97, 116; *Lombard North Central plc v Butterworth* [1987] QB 527 (CA) 542-43; *O'Dea v Allstates Leasing System (WA) Pty Ltd* (1983) 152 CLR 359 (HCA) 367; *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 184-85, 197, 205, 211.

115 *Bridge v Campbell Discount Co Ltd* [1962] AC 600 (HL) 624; *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 185.

116 *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 211.

First, undoubtedly, the jurisprudence dictates that a clause activated by a breach of contract provides the paradigm scenario,[117] or 'core case', where the court's review jurisdiction arises. Despite the *Andrews* decision, in Australia the 'general rule' is that only a breach is required.[118] Why does breach matter? Chapter 8 explores this issue in detail. In summary, it concludes that a breach of a duty, such as a breach of contract, is a normatively significant event because it violates corrective justice norms. It does so by disturbing the parties' initial equilibrium. As a result, corrective justice requires a particular response to such an event, which will restore that equilibrium. Further, *only* corrective justice can determine the acceptable response to a violation of its norms. This point is fundamental. Chapter 8 will return to it. For now, suffice it to say that breach of contract provides the 'core case' for engaging the Penalty Rule because it is an event that is normatively significant, and one that is crucial to the rule's proper justification.

Second, given that breach is fundamentally important, Anglo-Australian private law should not *abolish* the requirement; rather, it should get serious about it and properly *respect* it. This means it is essential that courts take a substantive approach to the jurisdictional question. Is it legitimate for courts to do so? On several occasions, the highest courts have held that decisions regarding the Penalty Rule jurisdiction turn on substance, not form.[119] Further, as regards the detriment requirement, the justification for the applying the rule to the transfer of proprietary rights is that this issue is one of substance, not form.[120] Accordingly, it follows that courts legitimately may take a substantive approach to the Penalty Rule jurisdiction.

Third, respecting the breach requirement means the courts must review clauses where a breach is not strictly the detriment-triggering event, but is the causally operative triggering event. A breach will be causally 'operative' if the direct detriment-triggering event cannot occur without a breach (that is, a breach is a necessary condition for that event to occur).

To an extent, Anglo-Australian courts already accept this principle. As discussed, where A's exercise of her right to terminate for breach generates B's obligation to pay an agreed damages sum, the clause is reviewable. In such cases, termination based upon a breach is basically tantamount to a breach. In other words, *breach is the causal trigger* for B's detriment, and the Penalty Rule applies. The rationale is clear: the length of any causal chain is unrelated

---

117 Text following n 20 in ch 1.

118 *Carter on Contract* [43-160] (service 37, April 2014).

119 eg *Bridge v Campbell Discount Co Ltd* [1962] AC 600 (HL) 624; *O'Dea v Allstates Leasing System (WA) Pty Ltd* (1983) 152 CLR 359 (HCA) 368, 398-99, 403.

120 *Wollondilly Shire Council v Picton Power Lines Pty Ltd* (1994) 33 NSWLR 551 (NSWCA) 555.

to the normative connection between B's conduct and the detriment he incurs. Where a breach lies at the head of the chain, and it forms the operative catalyst for the causal events that follow, then the length of the casual chain is irrelevant. In other words, the breach may be a direct cause, or a remote cause. In this thesis, a 'causal breach' refers a breach that is the causally operative trigger for B's detriment.

Finally, respecting the breach requirement means the courts must not allow a party to evade that requirement.

To a degree, English courts already accept this principle. As outlined, English courts have held that, in principle, the Penalty Rule will catch situations where one party adopts a drafting avoidance technique—such as drafting a clause that on its face appears to be option clause or appears to be an incentive payment clause—to create a 'disguised' agreed damages clause that seeks to circumvent the jurisdiction. In such cases, *a substantive breach is the trigger* for B's detriment, and the Penalty Rule is engaged. In this thesis, a 'default' refers to an actual or substantive breach of contract collectively.

Long-standing Australian authority also suggests that 'disguised' agreed damages clauses will not escape review. In *Bysouth*, the Full Court of the Victorian Supreme Court rejected an attempt to disguise a penalty clause in some other form so as to evade the Penalty Rule. Irvine CJ held that the rule applied to the relevant clause despite the fact that the words used constituted 'an ingenious paraphrase intended to avoid the consequences that would follow had the stipulation been put in more direct form'.[121]

This analysis leads to the following question: what is the proper justification for the review power applying to cases that do not involve a formal rights infringement? It is *anti-stultification*. To stultify is 'to make a fool of' or 'to make nonsense of'.[122] In this instance, if A adopts an avoidance technique to circumvent the jurisdiction, failing to review the clause would make nonsense of the law's commitment to reviewing agreed damages clauses with a view to invalidating penalty clauses. Consequently, the Penalty Rule would be stultified. Such circumvention is therefore impermissible. Preventing stultification gives full effect to the breach requirement. It ensures that courts review *all* clauses that truthfully can be said to involve a breach. As the English Court of Appeal has stated, the Penalty Rule jurisdiction 'applies to clauses operating on breach, whatever be the nature of the breach'.[123]

---

121 *Bysouth v Shire of Blackburn and Mitcham (No 2)* [1928] VLR 562 (VSCFC) 574.

122 Peter Birks, 'Recovering Value Transferred under an Illegal Contract' (2000) 1 Theo Inq L 155, 160.

123 *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [52].

Accepting the anti-stultification rationale creates consistency within the Penalty Rule. When applying the tests of validity, the parties' use of the phrase 'penalty' or 'liquidated damages' is relevant but not conclusive. Labels are not determinative. If they were, one party would be able to include a substantive penalty clause in her contract, but formally characterise the term as a 'liquidated damages clause' and ensure that a court upheld it. This would make nonsense of the private law's stance against penalty clauses.

This rationale is powerful. The Law Commission has noted that the greatest problem in extending the Penalty Rule review power is this: how does one define the jurisdictional scope in a way that does not subject to review *every* term involving an obligation to pay?[124] The anti-stultification rationale of the jurisdictional 'extension' solves this dilemma, for it provides a clear limit on the court's proper review jurisdiction. In short, the (extended) rule will only catch a clause if one party is trying to achieve underhandedly what she could not achieve openly.

II Revisiting the case law

The English and Australian approaches to the jurisdictional issue are aligned on some points, and appear disparate on others. However, the principled approach can reconcile them based on the view that the Penalty Rule catches clauses whereby a causal or substantive breach is the trigger for B's detriment. This view best explains the present law. It is now necessary to re-examine the case law with that understanding in mind.

Consider first cases where a causal breach leads to B's detriment. Clearly, Anglo-Australian courts accept that the Penalty Rule is engaged where the event activating the clause is not a breach itself, but an event directly conditioned upon a breach. A prime example is A's exercise of her right to terminate for B's breach. Some clauses, however, require an extra step. For example, some putative agreed damages clauses condition the detriment upon A terminating for breach *plus* performing an additional act, such as A exercising a contractual 'buy-back' option. At present, it is an open question whether a court will view the events as tantamount to a breach, and thus see the breach as the causally operative trigger upon which B incurs the detriment.[125] The principled approach suggests that the review power extends to cases where a breach lies at the head of the chain, and forms the operative catalyst for the causal events that follow. The latest word on the issue supports this approach:

---

124 Law Commission, *Penalty Clauses and Forfeiture of Monies Paid* (Law Com WP No 61, 1975) [22].

125 *Ringrow Pty Ltd v BP Australia Pty Ltd* [2005] HCA 71, (2005) 224 CLR 656 [21], [36] (issue left open); *Jervis v Harris* [1996] Ch 195 (CA) 204 (against).

> Nor do I consider that the law of penalties applies only if the wrongdoing is a necessary and sufficient cause of the charge being payable … The fact that a payment is levied only if the breach is committed in particular circumstances or if other conditions are satisfied does not in itself mean that it is not paid on breach.[126]

Turning now to cases where a detriment arises upon a substantive breach, English law is seemingly clear, whereas Australian law is more difficult.

In England, properly understood, the case law has long accepted that the Penalty Rule jurisdiction extends to substantive breaches, and acknowledged the anti-stultification rationale. As to the former, in *Clydebank Engineering and Shipbuilding Co Ltd v Don Jose Ramos Yzquierdo y Castaneda*, Lord Halsbury held that courts 'must proceed according to what is the real nature of the transaction'.[127] As to the latter, in *Cooden Engineering Co v Stanford*, Hodson LJ said he was unable to accept a contention that a drafter 'can always draw [her] document in such a way as to defeat the [rule]'.[128] In sum, the principled approach provides the most plausible, coherent and appealing account of English law as it stands.

In Australia, one question is pressing: does the *Andrews* judgment support a substantive approach, and an anti-stultification rationale?

Three features of the *Andrews* decision support the view that, as best interpreted, the Court applied a substance test.

To start, in defining the penalty clause concept, the High Court cited Deane J's judgment in *Acron Pacific Ltd v Offshore Oil NL*, in which his Honour had held that the Penalty Rule applies in cases of substantive breaches.[129]

Next, the High Court praised Brereton J's judgment in *Interstar*. The Court endorsed the judge's view that the Penalty Rule is engaged where a clause is activated by a failure of a condition that is 'in substance equivalent to a promise that the condition will be satisfied'.[130] It also referred to and cited with approval the passage of the *Interstar* judgment containing the Responsibility Test, which is a test focusing on the parties' *substantive* understanding of B's

---

[126] *Office of Fair Trading v Abbey National plc* [2008] EWHC 2325 (Com Ct), [2009] 1 All ER (Comm) 717 [19].

[127] [1905] AC 6 (HL) 9.

[128] [1953] 1 QB 86 (CA) 116.

[129] (1985) 157 CLR 514 (HCA) 520.

[130] *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [67] (footnote omitted).

responsibilities.[131] Notably, judges and jurists have interpreted the High Court in *Andrews* as endorsing Brereton J's reasoning, and a substantive approach. For example, it has been held that the extended jurisdictional test requires a court to identify a stipulation that is substantively equivalent to a promise that an event will occur, or will not occur.[132]

Lastly, as previously noted, in *Andrews* the Court held that the review jurisdiction is not invoked if the loss A suffers from the primary stipulation failing cannot be evaluated or assessed in money terms. Interpreted in the abstract, this proposition was shown to be problematic. However, this statement makes sense if the High Court is understood to have extended the jurisdiction to cover stipulations that are substantively equivalent to promissory obligations. That is, the loss must be 'capable of being financially assessed' in the following sense: the event must be sufficiently similar to a breach in that the occurrence or failure of the event must be capable of causing what accurately can be described as a 'loss', requiring a compensatory response. As Carter and others have noted, a court can only assess the loss that a failure of condition causes if the condition 'was the subject of some sort of promise'.[133] Further, if an agreed sum is payable upon 'an event closely allied to breach', a court can sensibly apply any comparative test of validity, as the clause's purpose will be either 'to pre-estimate the claimant's loss caused by the "event" or to punish the defendant …'.[134] Compare the case where the condition generates an outcome that correctly cannot be described as a 'loss'. In that instance, a court cannot provide a *compensatory* remedy, for as there is nothing to repair a court has no 'handle to grant relief';[135] and a court cannot apply any comparative test of validity that depends upon pre-assessing A's anticipated *loss*, for a different concept is in play.

As for the anti-stultification rationale, the High Court implicitly supported the view that a party cannot avoid the jurisdiction through clever drafting. Two reasons follow.

On the one hand, in *Interstar,* Brereton J rejected the view that the Penalty Rule 'could always be evaded by the drafting of lists of event of default … without including a contractual

131 *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [49]. As to the Responsibility Test, see text to n 47.

132 *Love v Brien* [2012] WASC 457 [62], affd [2013] WASCA 280. See also Kanaga Dharmananda and Leon Firios, 'Penalties Arising Without Breach: The Australian Apogee of Orthodoxy' [2013] LMCLQ 145, 148.

133 JW Carter and others, 'Contractual Penalties: Resurrecting the Equitable Jurisdiction' (2013) 30 JCL 99, 120.

134 Andrew Burrows, *Remedies for Torts and Breach of Contract* (3rd edn, OUP 2004) 453.

135 *Peachy v Duke of Somerset* (1720) 1 Str 447, 453; 93 ER 626, 630.

promise that those events would not occur'.[136] In *Andrews,* the High Court implicitly endorsed this analysis.[137]

On the other hand, in *Andrews*, the successful applicants argued that 'despite the form of the [relevant fee], in substance it [was] a disguised penalty'.[138] The Court did not disapprove of that proposition. As discussed, a 'disguised' penalty clause refers to an invalid clause drafted in an attempt to evade—and hence stultify—the jurisdictional rule. Accordingly, the *Andrews* judgment is consistent with the view that the review power will catch such clauses.

This argument raises one issue. It is the High Court's analysis of *Metro-Goldwyn-Mayer Ltd*. As outlined, on one reading the Court held that *true* option clauses are properly unreviewable, and then used that case to illustrate the point. In *Andrews*, the correct principle was identified. But the principle was not applied correctly.

On the facts of *Metro-Goldwyn-Mayer Ltd*, the 'optional' hiring fee was four times the usual hiring fee. The facts were therefore comparable to those of *Interfoto*. On two occasions, the English Court of Appeal has (rightly) held that the latter case involved a 'disguised' penalty clause. It is true that the facts of *Interfoto* were stronger, as the 'holding fee' represented an amount 10 times higher than the fee other comparable photographic libraries charged for a similar service. However, three aspects of *Metro-Goldwyn-Mayer Ltd* are noteworthy.

First, the entire court agreed upon the nature of the increased 'price'. It was variously described as 'extravagant', and as involving an 'apparent enormity of … multiplication'.[139] In Jacobs JA's words, the contract may well have intended 'that such an additional showing should be strongly discouraged. For this reason a very large hiring fee compared with the original hiring fee [was] provided.'[140]

Second, the distributor led no evidence of the special circumstances justifying why the further 'right' or 'benefit' was priced at such a greater 'fee'. Only Holmes JA sought to justify the 'fee' increase, stating that 'the economic consequences of each exhibition [are] likely … to return a greater profit to the exhibitor'.[141] This reasoning is unconvincing. To begin with, it is

---

[136] *Integral Home Loans Pty Ltd v Interstar Wholesale Finance Pty Ltd* [2007] NSWSC 406, (2007) 2 BFRA 23 [73], revd [2008] NSWCA 310, (2008) 257 ALR 292.

[137] *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [49].

[138] *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [28].

[139] *Metro-Goldwyn-Mayer Ltd v Greenham* [1966] 2 NSWR 717 (NSWCA) 720, 726.

[140] *Metro-Goldwyn-Mayer Ltd v Greenham* [1966] 2 NSWR 717 (NSWCA) 723.

[141] *Metro-Goldwyn-Mayer Ltd v Greenham* [1966] 2 NSWR 717 (NSWCA) 727.

implausible. Also, if it was correct, the fee for each additional screening should have gradually increased.

Finally, a further clause of the agreement clarified that if the unauthorised showing was due to circumstances beyond the exhibitor's control, then the normal hiring fee was payable; if it occurred by reason of the exhibitor's 'negligence or default',[142] the multiple fee was payable. Accordingly, the exhibitor could only exercise his option, and pay the increased price, where he controlled whether the rescreening occurred and was at fault. As Chapter 10 explains, in these circumstances, properly analysed the clause was a classic example of an obligation 'dressed up' as an option. It should have been reviewable. But that does not detract from the correctness of the High Court's general principle.

It follows that the *Andrews* judgment is best understood as propounding a substance test encapsulating the anti-stultification rationale.

The principled approach, then, presents a model for comprehending the breach requirement that is perspicuous and reasoned. It shows that this limb of the jurisdictional test in Anglo-Australian law can be reduced to a single question: is an actual breach the causally operative trigger for B's detriment, or is an actual or substantive breach the trigger for B's detriment? Put differently, is B's detriment conditioned upon a causal breach or a default? A positive answer will satisfy the second limb of the jurisdictional test.

4 Conclusion

This chapter has illustrated that, properly understood, the Penalty Rule jurisdictional test contains two requirements: B must suffer a detriment, and B's breach of contract must trigger the detriment.

For Anglo-Australian courts to obtain jurisdiction, B must incur a relevant detriment as a result of the clause. The detriment must be that B comes under an obligation to pay money to A; to transfer a personal or proprietary right to A; or to extinguish a right held against A that obligates A to transfer money or property to B.

Traditionally, the breach requirement has limited the court's review power. Until recently, generally that restriction was regarded as irreproachable. As *Andrews* demonstrates, this is not the case. However, it is rational principle, not history, which demands this revision. Accordingly, English courts' limited review power must be interpretively extended beyond

[142] *Metro-Goldwyn-Mayer Ltd v Greenham* [1966] 2 NSWR 717 (NSWCA) 719.

cases involving the strict violation of a contractual duty. Conversely, Australian courts' extended jurisdiction must be interpretively limited. Undoubtedly, the jurisprudence dictates that breach provides the 'core case'. However, the review jurisdiction also properly includes agreed damages clauses that are triggered by a causal or substantive breach. As such, the review power properly extends to clauses activated by a default or a causal breach.

5 VALIDITY

In *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd*,[1] Lord Dunedin outlined several tests for identifying whether a clause is valid under the Penalty Rule. His Lordship's formulation of the tests of validity sought to reconcile centuries of jurisprudence. The rule's longevity evidences his success in doing so.

This chapter offers a systematic and principled examination of each proposed test of validity, and the relationship between particular tests. It pays close attention to two key questions. Which of these tests, if any, is the primary or sole test applied? Does each test still work? It is argued that, as regards the tests of validity, Anglo-Australian law has worked itself pure.

The chapter will then analyse three problematic issues. First, it examines recent case law suggesting that the test of validity is subject to a qualifier. Specifically, it considers whether a proper 'commercial justification' can save a prima facie penalty clause. Second, the chapter examines the controversial question of the proper comparator for any test of validity to adopt. Put differently, when evaluating an agreed compensatory damages clause, is the correct benchmark for assessment anticipated *actual* loss, or anticipated *legally-recoverable* loss? Finally, the chapter clarifies the ambiguity regarding the proper assessment date, and how a court is correctly to apply a test operating as at that date. A rational approach to these problems reveals that the first is one where the stakes are high, for a wrong answer threatens the doctrine's very existence, but the solution is clear; the second 'problem' is an illusion; and though the final problem is easily solved, the lesson learnt is critical.

1 The potential tests

Generally speaking, the Penalty Rule test of validity seeks to identify the 'additional detriment'[2] that a penalty clause imposes upon B. Presently, there are five tests seemingly available to a court to help it to decide whether a clause is penal. As Chapter 3 discussed, these are the Estimation Test; the Extravagance Test; the Greater Sum Test; the Indiscriminate Cause Test; and the Impossibility Test. Each test is considered in sequence.

---

[1] [1915] AC 79 (HL).

[2] *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [10].

A The Estimation Test

I The primary or sole test

The Estimation Test refers to the test that determines whether an agreed damages clause is valid based on whether the relevant detriment constitutes a genuine pre-estimate of loss. Is the Estimation Test the primary, or sole, test of validity?

One must begin the analysis with *Dunlop*. In that case, Lord Dunedin held that Estimation Test identifies the 'essence' of liquidated damages, whereas the other tests were only used in construing the contract. This suggests that the Estimation Test is primary. Conversely, his Lordship also held that the other tests 'may prove … conclusive'. This suggests that the remaining tests are more than merely aids. The case, then, is inconclusive.

McGregor takes the traditional view that the Estimation Test is primary: it 'forms the basis of the distinction' between liquidated damages clauses and penalty clauses, such that the remaining tests 'stem from it and are subordinate to it'.[3] A case exemplifying this view is *Robophone Facilities Ltd v Blank*, where Diplock LJ held that a court will not enforce an agreed damages clause if the stipulated sum 'was not a genuine estimate of the loss likely to be sustained' upon breach.[4]

Momentarily it will be illustrated that recent case law in England and Australia has moved from this position. That said, on their face, some recent English cases support the conventional view. For example, in *Lansat Shipping Co Ltd v Glencore Grain BV (The Paragon)*, the English Court of Appeal held that an agreed damages clause is 'unenforceable unless the payment stipulated for is a genuine pre-estimate of the amount of loss caused by the breach'.[5] Also, in *Makdessi v Cavendish Square Holdings BV*, the Court of Appeal considered that the principles arising from the case law provided merely 'guidelines for the purposes of determining whether the clause is a genuine pre-estimate'.[6]

For three reasons, these judicial comments are not determinative. To start, in *Lansat*, the matter of whether any particular test was paramount was not in issue; hence, the judgment did not authoritatively decide the point. Next, in *Lansat*, the Court of Appeal seemingly failed

---

[3] Harvey McGregor, *McGregor on Damages* (19th edn, Sweet and Maxwell 2014) [15-028]. See similarly Chris Rossiter, *Penalties and Forfeiture* (Lawbook Company 1992) 32.

[4] [1966] 1 WLR 1428 (CA) 1446.

[5] [2009] EWCA Civ 855, [2010] 1 All ER (Comm) 459 [31]. See also [33].

[6] *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [71].

to appreciate the distinction between the Estimation and Extravagance Tests, treating them as interchangeable.[7] As illustrated below, they are not. Finally, in *Makdessi*, the Court of Appeal meant what is said, but it did not say what it meant. We return to this point shortly.

II Interpretation

The Estimation Test refers to a 'genuine' pre-estimate of loss. Courts have yet categorically to define what the term means. Does it mean 'substantially accurate', 'honest' or 'reasonable'? Each interpretation is examined in order.

(1) Substantially accurate

In *Dunlop*, Lord Parmoor spoke of the need for an 'accurate pre-estimate of damage'.[8] Building on this, in the past a strand of English and Australian authorities seemingly supported a 'substantially accurate' interpretation of the expression 'genuine'. These cases mostly involved chattel leases[9] and hire purchase[10] agreements. These cases demanded near exactitude between the agreed compensatory damages amount and the general compensatory damages that a court would award for the particular breach. For example, in *Citicorp Australia Ltd v Hendry*, Priestly JA held that a 'genuine' pre-estimate required that the agreed damages clause 'leave in the lessor's hands … a sum approximating that which fulfilment of the contract would have yielded'.[11] According to the Victorian Law Reform Commission, the *Citicorp* judgment required a 'standard of accuracy' that approached 'near precision'.[12]

Some writers continue to favour a substantially accurate interpretation, at least in some contexts.[13] However, recent authorities have undermined this view. In *Murray v Leisureplay plc*, Arden LJ held that the parties 'do not have to make an accurate assessment of the damages that would have been awarded at common law'; and Buxton LJ held that the phrase 'genuine' does not mean that the agreed damages sum 'must be in fact an accurate statement of the

---

[7] The Court of Appeal adopted the trial judge's reasoning: *Lansat Shipping Co Ltd v Glencore Grain BV (The Paragon)* [2009] EWCA Civ 855, [2010] 1 All ER (Comm) 459 [31]. The trial judge, in turn, described a genuine pre-estimate of loss as one which was 'not extravagant': *Lansat Shipping Co Ltd v Glencore Grain BV (The Paragon)* [2009] EWHC 551 (Com Ct), [2009] 2 All ER (Comm) 12 [22].

[8] *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd* [1915] AC 79 (HL) 102, 104.

[9] eg *Citicorp Australia Ltd v Hendry* (1985) 4 NSWLR 1 (NSWCA).

[10] eg *Bridge v Campbell Discount Co Ltd* [1962] AC 600 (HL).

[11] (1985) 4 NSWLR 1 (NSWCA) 34.

[12] Victorian Law Reform Commission, *Liquidated Damages and Penalties* (VLRC DP No 10, 1988) [32].

[13] eg JW Carter, *Carter's Breach of Contract* (LexisNexis Butterworths 2011) [13-09].

loss'.[14] Further, consider *Zachariadis v Allforks Australia Pty Ltd*.[15] The facts involved a hiring agreement. Counsel submitted that traditionally courts had applied a 'substantial accuracy' test in determining whether agreed damages clauses in chattel leases and hire purchase agreements imposed a penalty, and that courts continued to do so. The Victorian Court of Appeal rejected the argument that a 'substantial accuracy' interpretation applies at all, even in the hire agreement context.[16]

As a matter of principle, this position must be correct. As Clarke J has extra-judicially argued, this interpretation renders problematic the compound phrase 'genuine pre-estimate'. On this view, one interprets the term 'genuine' as referring to 'substantially accurate'. If so, this will distort the meaning of the term 'pre-estimate'. In his Honour's words, 'The essence of an estimate is approximation. An estimate represents an assessment of the likely damages and [can] hardly be expected to be completely accurate.'[17]

Simply put, the estimate need not be accurate.

(2) Honest

Many scholars argue that term 'genuine' means 'good faith' or 'honest'. For example, McKendrick contends that what is at issue is the parties' *bona fides*.[18] The writer relies upon leading cases, such as *Dunlop*, which refer to a '*bona fide*' estimate.[19]

The contemporary authorities have expressly rejected this interpretation. In *Murray*, Buxton LJ's view was clear, his Lordship holding that 'genuine' does not mean 'honest'.[20] In *Makdessi*, the Court of Appeal confirmed this, declaring that the expression 'genuine' does not turn upon the 'honesty of the party or parties who made the pre-estimate. The test is primarily an objective one …'.[21]

---

[14] [2005] EWCA Civ 963, [2005] IRLR 946 [51], [111].

[15] [2009] VSCA 258, (2009) 26 VR 47.

[16] *Zachariadis v Allforks Australia Pty Ltd* [2009] VSCA 258, (2009) 26 VR 47 [91], [140]-[142].

[17] MJR Clarke, 'Penalties, Forfeiture and Dilution' [1989] AMPLA Ybk 1, 3.

[18] Ewan McKendrick, 'Liquidated Damages and Related Clauses in Claims involving Chattels' in Norman Palmer and Ewan McKendrick (eds), *Interests in Goods* (2nd edn, LLP 1998) 957.

[19] *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd* [1915] AC 79 (HL) 97; *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 181.

[20] *Murray v Leisureplay plc* [2005] EWCA Civ 963, [2005] IRLR 946 [111].

[21] *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [58] (footnote omitted).

In principle, the English courts' rejection of this view must be right. In this context, 'honesty' refers to the parties' subjective mental state. Accordingly, if the parties honestly attempted to predetermine the loss, the clause should be valid in any event. Conversely, all the tests of validity, including the Estimation Test, require an assessment that is either wholly objective[22] or primarily objective taking account of subjective factors.[23] Consequently, the requisite inquiry 'necessitates an objective assessment of the possible loss that might be incurred, not the subjective views or calculations of the parties prior to entering into the contract'.[24] Thus, the fact the parties' attempt was honest will not ensure that the clause is upheld.

(3) Reasonable

A further interpretation of the expression 'genuine' suggests that it means 'reasonable'. There is a line of English judicial support for this interpretation.[25] For instance, in *Liberty Life Assurance Co Ltd v Sheikh*, Kerr LJ interpreted the Estimation Test as requiring a 'reasonable pre-estimate' of loss.[26] Further, in *Alfred McAlpine Capital Projects Ltd v Tilebox Ltd*, Jackson J supported the view that courts require a reasonable pre-estimate, and further held that the estimate is only unreasonable if a 'substantial discrepancy' exists between the agreed compensatory damages and the anticipated loss.[27] Recently, the English Court of Appeal has approved this analysis.[28]

Peel supports a reasonableness interpretation. The writer argues that the Privy Council's decision in *Workers Trust and Merchant Bank Ltd v Dojap Investments Ltd*[29] illustrates that the Penalty Rule applies to deposit clauses, and he relevantly interprets that case as holding that a deposit clause is penal when it is does not represent a 'reasonable pre-estimate of the loss'.[30] For two reasons, this argument is difficult.

---

[22] *Murray v Leisureplay plc* [2005] EWCA Civ 963, [2005] IRLR 946 [55]; *State of Tasmania v Leighton Contractors Pty Ltd* [2005] TASSC 133, (2005) 15 Tas R 243 [31(5)].

[23] *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [58] (footnote omitted).

[24] *Paciocco v Australia and New Zealand Banking Group Ltd* [2014] FCA 35, (2014) 309 ALR 249 [41]. See also [126].

[25] But see *O'Dea v Allstates Leasing System (WA) Pty Ltd* (1983) 152 CLR 359 (HCA) 398, 400.

[26] The Times, 25 June 1985 (CA).

[27] [2005] EWHC 281 (TCC), [2005] BLR 271 [48(1)].

[28] *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [81], [93], [105].

[29] [1993] AC 573 (PC).

[30] Edwin Peel, *The Law of Contract* (13th edn, Sweet and Maxwell 2011) [20-149].

On the one hand, deposits clauses are in fact not subject to the Penalty Rule. This is because in the case of a deposit a purchaser does not 'forfeit' (that is, give up) any right upon breach, such that the purchaser suffers no relevant detriment sufficient to engage the review jurisdiction. It follows that any interpretation of the meaning of 'genuine' derived from a deposit case, such as *Workers Trust*, is necessarily problematic.

On the other hand, on its own terms, the analysis arguably does not work. In *Workers Trust*, Lord Browne-Wilkinson held that a deposit sum might be 'reasonable' even if it 'bears no reference to the [vendor's] anticipated loss'.[31] It necessarily follows that a deposit sum may be 'reasonable' even if it *not* a reasonable pre-estimate of loss. Accordingly, one cannot interpret 'reasonable' as a synonym for 'genuine' in the Estimation Test.

It is important to note the argument made here. The argument is *not* that a reasonableness interpretation is incorrect. That is yet to be established. The argument *is* that *Workers Trust* cannot support such an interpretation.

In sum, the true meaning of the term 'genuine' in the Estimation Test remains doubtful. Regardless, on any interpretation of that expression, the Estimation Test leads to incoherence, and lacks fit. We turn to these issues now.

III Coherence

Some judges and jurists see the Estimation Test and Extravagance Tests as coterminous. For example, in *Philips Hong Kong Ltd v A-G of Hong Kong*, the Privy Council held that a sum will be a 'genuine pre-estimate of … loss' provided that it 'is not extravagant'.[32] This is incorrect. The two tests are conceptually distinct. The fact the two tests are separate means they may produce differing results when applied to the same set of facts.

On all putative interpretations of the expression 'genuine', the Estimation Test and the Extravagance Test are not coterminous, and potentially may be inconsistent.

First, assume 'genuine' means substantially accurate. An inaccurate pre-estimate of loss is not *necessarily* extravagant. The Extravagance Test inherently sanctions a large discrepancy between the agreed compensatory damages amount and the anticipated loss, for the parties are allowed a 'generous margin'.[33] On this view, the Estimation Test allows for no such

---

[31] *Workers Trust and Merchant Bank Ltd v Dojap Investments Ltd* [1993] AC 573 (PC) 578.

[32] (1993) 61 BLR 41 (PC) 59.

[33] *Murray v Leisureplay plc* [2005] EWCA Civ 963, [2005] IRLR 946 [43]; *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [75(iv)].

discrepancy. In principle, then, many agreed damages amounts would pass the Extravagance Test, but fail the Estimation Test.

Second, suppose 'genuine' means good faith. If so, the clause may fail the Estimation Test, if the parties fail honestly to estimate the anticipated loss that would likely arise from the breach. Despite this, the agreed compensatory damages clause may pass the Extravagance Test, for the estimated sum may be in any event not wholly disproportionate to the anticipated loss.

Finally, consider the position where a reasonable interpretation applies. In ordinary parlance, when used to describe a sum the expression 'reasonable' means 'moderate'.[34] This suggests that a 'reasonable pre-estimate' test sanctions only a moderate divergence between the agreed damages sum and the anticipated loss. Conversely, the Extravagance Test does what the label says, and sanctions a much greater discrepancy.

However, in *Alfred McAlpine*, Jackson J attempted to reconcile the Extravagance Test and a 'reasonable pre-estimate' test, holding that an estimate will be reasonable even if not substantially accurate, and that an estimate is only unreasonable if a 'substantial discrepancy' exists between the agreed compensatory damages amount and the anticipated loss.[35] Recently, the English Court of Appeal endorsed this analysis.[36] This attempted reconciliation is laudable. It fails, however, for two reasons.

For one, this 'reconciliation' involves an abuse of language. As stated, reasonable means moderate. But on his Lordship's interpretation, a sum which bears (possibly very) little relation to the anticipated loss remains 'reasonable' provided that the disproportionality is something *less than substantial*. This approach is casuistic. Take the facts of *Alfred McAlpine* itself. The construction contract between Tilebox and Alfred McAlpine contained an agreed compensatory damages clause. That provision stated that if Alfred McAlpine failed to complete the construction works on time, it would be obliged to pay Tilebox £45,000 per week for the period of its delay. Based on one set of hypothetical calculations, Jackson J estimated that Tilebox would suffer anticipated direct losses of £7,500 per week.[37] If this is

---

[34] Oxford Dictionaries <www.oxforddictionaries.com/definition/english/reasonable?q=reasonable> accessed 12 January 2015.

[35] *Alfred McAlpine Capital Projects Ltd v Tilebox Ltd* [2005] EWHC 281 (TCC), [2005] BLR 271 [48(1)].

[36] *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [93].

[37] *Alfred McAlpine Capital Projects Ltd v Tilebox Ltd* [2005] EWHC 281 (TCC), [2005] BLR 271 [95(1)] (comparison with the 'middle of the range'), [96].

correct, then the agreed damages were precisely *six times* the anticipated losses. Can this this discrepancy properly be described as 'moderate'?

For another, assume that, properly interpreted, 'unreasonable' means 'substantial discrepancy'. If so, then there is little discernable difference between this standard and one of 'extravagance'. This is why, in *Makdessi*, the Court of Appeal asked whether the relevant agreed damages clauses were 'extravagant and unreasonable (in the sense employed by Jackson J …)'.[38] The Court was (appropriately) adopting 'unreasonable' as a synonym for 'extravagant'. Accordingly, Jackson J's 'reasonable pre-estimate' test is not a reinterpreted Estimation Test at all; rather, it is merely the Extravagance Test in disguise. It is therefore redundant, and liable to confuse.

The fact that the Estimation Test and Extravagance Test are inconsistent leads to two conclusions. On the one hand, it results in a test of validity that is not even weakly coherent, for in some cases it is internally inconsistent. An incoherent law is less justifiable. On the other hand, if judges fail to recognise the distinction between the two tests, they may unknowingly apply substantively different tests to determine cases displaying similar fact patterns. This will violate the uniformity principle, which requires similar treatment of like cases. These problems are resolved if, properly understood, one or the other test forms the sole Penalty Rule test of validity.

IV Fit

The Estimation Test is under-inclusive, as it cannot account for a significant segment of the jurisprudence. Put differently, it lacks fit.

In *Cellulose Acetate Silk Co Ltd v Widnes Foundry (1925) Ltd*,[39] a contractor agreed to build an acetone recovery plant for a principal. The relevant clause provided that, if the contractor delayed, it would pay to the principal a 'penalty' of £20 for each week of delay.[40] When contracting, both parties knew a delay was likely to cause loss in excess of £20 per week: that is, the parties deliberately 'underestimated' the agreed damages amount. The contractor completed 30 weeks late. The principal suffered losses of £5,850. The contractors sued for the contract price; the principal argued the clause was penal. The House of Lords held that the clause was valid. The principal was only entitled to £600.

---

[38] *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [105].

[39] [1933] AC 20 (HL).

[40] *Cellulose Acetate Silk Co Ltd v Widnes Foundry (1925) Ltd* [1933] AC 20 (HL) 22.

Some writers have purported to justify *Cellulose* on the basis that the facts actually involved a limitation clause.[41] Like a limitation clause, the relevant clause restricted recovery to an amount less than the principal's incurred loss. However, two reasons suggest that the clause was not a limitation clause. First, unlike a limitation clause, the *Cellulose* clause *set* the contractor's liability at an exact sum it had to pay, even if the principal's incurred loss was *less* than £20 per week. Second, as liability may range, limitation clauses require proof of incurred loss. However, the amount payable under the *Cellulose* clause was recoverable without proof of loss.

The case is problematic because, however 'genuine' is defined, the outcome appears incorrect. Consider each potential definition in order.

To start, the pre-estimate was not substantially accurate. In *Cellulose*, the House of Lords acknowledged that 'it must have been obvious to both the parties that the [incurred loss] would be much more than £20 a week …'.[42]

Next, the pre-estimate was not honest. The parties were not trying to estimate the loss: they were engaged in a wholly different exercise of agreeing a value known in any event to be less than the expected loss. Put differently, the parties were not honestly *estimating* the loss; they were honestly *agreeing or setting* a figure. The latter, however, cannot satisfy an estimation-based test, for no estimation has taken place at all.

Lastly, the estimate could not be felicitously described as reasonable, for the parties knew the owner most likely would be left heavily undercompensated. Thus, according to the trial judge, the owner's incurred loss was within the parties' reasonable contemplation, and so they could have agreed 'a reasonable pre-estimate of [the] probable loss', but they did not; therefore, the sum was not a genuine pre-estimate.[43]

Accordingly, the House of Lords acknowledged that the clause failed the Estimation Test.[44] Leading commentators agree with this conclusion.[45]

Since *Cellulose,* both English[46] and Australian[47] courts have confirmed that a derisory agreed compensatory damages clause cannot be penal.[48] A striking example is as follows. In

---

[41] eg Michael Bridge, *The Sale of Goods* (3rd edn, OUP 2014) [12.12].

[42] *Cellulose Acetate Silk Co Ltd v Widnes Foundry (1925) Ltd* [1933] AC 20 (HL) 25-26.

[43] *Widnes Foundry (1925) Ltd v Cellulose Acetate Silk Co Ltd* [1931] 2 KB 393 (CA) 401, affd [1933] AC 20 (HL).

[44] *Cellulose Acetate Silk Co Ltd v Widnes Foundry (1925) Ltd* [1933] AC 20 (HL) 25.

[45] eg Harvey McGregor, *McGregor on Damages* (19th edn, Sweet and Maxwell 2014) [15-091]-[15-092].

[46] *Bath and North East Somerset District Council v Mowlem plc* [2004] EWCA Civ 115, [2004] BLR 153 [14].

Temloc Ltd v Errill Properties Ltd, a principal engaged a contractor to build a shopping centre development.[49] The relevant clause provided that if the contractor failed to complete the works on time, it would pay agreed damages for its delay at the rate stated in the Appendix. The Appendix stated that the 'rate' as '£Nil'. The contractor completed late. The principal argued that, properly construed, the relevant 'clause' was to be disregarded, such that the contract did not include an agreed damages clause. Consequently, the principal was entitled to sue for general compensatory damages. The contractor argued that the clause was valid, but the liquidated compensatory damages were simply nil. The English Court of Appeal held that the relevant clause constituted an 'effective' liquidated damages clause for a nil amount, and that the contractor was not liable to pay to the principal any damages—liquidated or general—for its late completion.[50] Relevantly, the stipulated sum was enforceable as liquidated damages, but it clearly did not represent a genuine pre-estimate of loss.

How can a clause fail the Estimation Test, and still be upheld? According to the New South Wales Supreme Court in *Multiplex Constructions Pty Ltd v Abgarus Pty Ltd*, this is because the 'true vice' of a penalty clause is that it provides for a result that '*exceeds* that which a genuine pre-estimate of [loss] would have yielded'.[51] English courts agree. For example, in *Cellulose*, Scrutton LJ found 'great difficulty in saying that an estimate less than the [anticipated] loss can ever be a penalty …'.[52] To talk of a penalty clause that is less than the anticipated loss involves 'a contradiction in terms'.[53] It follows that the rule operates in one direction only. A 'one way only' rule suggests a concern with preventing *excessive* recovery, rather than assuring an accurate or reasonable pre-estimation.

In conclusion, the Penalty Rule's primary test of validity must be certain, consistent and account for the majority of the case law. Conversely, the Estimation Test's true meaning is ambiguous; it is incompatible with the Extravagance Test; and it is under-inclusive. It follows that either the Penalty Rule's primary test of validity flawed, or the Estimation Test is not the primary test.

---

47 *Multiplex Constructions Pty Ltd v Abgarus Pty Ltd* (1992) 33 NSWLR 504 (NSWSC) 519.

48 Unlike in the paradigm scenario, in cases of underestimated agreed damages clauses it is A (the recipient), not B (the payer or transferee), who alleges that the clause is penal.

49 (1987) 39 BLR 30 (CA).

50 *Temloc Ltd v Errill Properties Ltd* (1987) 39 BLR 30 (CA) 38, 39-40.

51 (1992) 33 NSWLR 504 (NSWSC) 519 (emphasis added).

52 *Widnes Foundry (1925) Ltd v Cellulose Acetate Silk Co Ltd* [1931] 2 KB 393 (CA) 407, affd [1933] AC 20 (HL).

53 JL Barton, 'Penalties and Damages' (1976) 92 LQR 20, 25.

B The Greater Sum Test

Under the Greater Sum Test, a clause is a penalty clause if the breach consists only in not paying a sum, and the sum stipulated in the clause is a greater sum.

According to Lord Dunedin, the Greater Sum Test is a 'true corollary' to the Extravagance Test. However, that statement cannot be correct. The two tests adopt a different yardstick for comparison, and apply a different standard for evaluation. On the one hand, the Greater Sum Test compares the agreed damages amount with the sum B must pay to A under a separate provision. The test necessarily requires absolute exactitude between the sum payable and the sum due. On the other hand, the Extravagance Test adopts as its base figure either A's anticipated actual loss, or her anticipated legally-recoverable loss. By definition, when applied to an obligation to pay agreed damages for failure to pay a debt, this test permits a (possibly substantial) discrepancy between the sum payable and the sum due. A corollary is a logical proposition that follows from one already proved. However, an agreed damages clause that fails the Greater Sum Test will not *necessarily* fall foul of the Extravagance Test. For example, assume an agreed compensatory damages clause requires B to pay to A £150 if he fails to pay to A £100 on the due date. This clause would fail the Greater Sum Test. However, 'a creditor's obligations may mean that [her] loss exceeds the sum due and the sum stipulated could very well [pass the Extravagance Test]'.[54]

In this case, however, time has wreaked the most havoc. In *Dunlop*, Lord Dunedin held that the Greater Sum Test was premised on the rule that compensatory damages were unrecoverable for the late payment of a debt.[55] Anglo-Australian law now holds that a claimant may recover interest as compensatory damages for the late payment of a debt.[56] So, suppose there exists an agreed compensatory damages clause for breach of an obligation to pay a debt. The clause includes a component for the creditor's anticipated loss of the use of money which she would be forced to pay away, or which would be withheld from her, as a result of the debtor's breach.[57] In either case, the agreed compensatory damages amount

---

54 Andrew Burrows, *Remedies for Torts and Breach of Contract* (3rd ed, OUP 2004) 442.

55 *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd* [1915] AC 79 (HL) 87.

56 *Hungerfords v Walker* (1989) 171 CLR 125 (HCA); *Sempra Metals Ltd v IRC* [2007] UKHL 34, [2008] 1 AC 561 [18], [100].

57 The agreed damages amount might thus include either the expected interest to be foregone in the form of a lost opportunity to invest the money elsewhere, or the expected interest to be incurred on funds borrowed to replace the money paid away or withheld: *Hungerfords v Walker* (1989) 171 CLR 125 (HCA) 142-44, 149, 152 (losses 'plainly foreseeable' and not too remote). See similarly *Sempra Metals Ltd v IRC* [2007] UKHL 34, [2008] 1 AC 561 [95].

legitimately would reflect the creditor's likely loss resulting from the late payment, despite being a sum greater than the unpaid debt. The clause, then, would be valid.

In summary, the Greater Sum Test no longer works. Rightly understood, it no longer forms part of the Penalty Rule.

C The Indiscriminate Cause Test

The Indiscriminate Cause Test holds that a presumption of penalty arises when a single lump sum is payable upon one or more of several events occurring, some of which may cause serious harm and others trivial harm. If B establishes that an agreed damages clause fails the test, the legal onus shifts to A to prove that the clause is valid.[58] In *Makdessi*, the English Court of Appeal confirmed that the Indiscriminate Cause Test remains alive and well.[59] The same applies in Australian law.[60]

The Indiscriminate Cause Test takes a 'weakest link'[61] approach. So, a court must assess the anticipated loss flowing from the least serious breach to which the clause applies. If the agreed damages amount is disproportionate to that expected loss, the weakest link snaps, and a presumption of penalty is raised. If the presumption is sustained, the clause is invalid, despite the fact it also applied to some or many breaches causing anticipated loss for which the agreed damages amount was proportionate, and hence not penal.[62] Accordingly, a clause that satisfies the tests of validity regarding *probable* breaches likely to cause substantial loss may be invalidated because a breach likely to cause trivial loss *theoretically* may have also invoked the term.

Perhaps due to this, Anglo-Australian courts regularly sidestep this rule. The case law contains a variety of 'avoidance techniques'. Three are as follows.

---

[58] JW Carter and Elisabeth Peden, 'A Good Faith Perspective on Liquidated Damages' (2007) 23 JCL 157, 166 cf *Ringrow Pty Ltd v BP Australia Pty Ltd* [2005] HCA 71, (2005) 224 CLR 656 [35].

[59] *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [73]-[74].

[60] *Fermiscan v James* [2009] NSWCA 355, (2009) 261 ALR 408 [149]-[152].

[61] *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd* [1915] AC 79 (HL) 89.

[62] *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd* [1915] AC 79 (HL) 102-103; *Pigram v A-G (NSW)* (1975) 132 CLR 216 (HCA) 221.

First, in some cases, courts construe an agreed compensatory damages clause as applying only to major breaches causing substantial loss, not minor breaches resulting in trivial loss.[63]

Second, in other cases, courts disregard remote *possible* breaches for which the agreed damages sum would greatly exceed the amount of A's anticipated loss. Rather, courts focus upon the likely loss that will arise from the *probable* breaches. So, in *Philips Hong Kong Ltd v A-G of Hong Kong*, the Privy Council held that an agreed damages clause is not presumptively penal merely because there might exist hypothetical situations in which the agreed damages amount was wholly out of proportion to the loss likely to be suffered.[64] In *Makdessi*, the Court of Appeal's approach was paradoxical: on the one hand, it confirmed that the Indiscriminate Cause Test remains relevant; on the other, it effectively dismissed the test by supporting this avoidance technique.[65]

Finally, a judge may interpret the agreed damages provision to ensure that the amount payable is proportionate to the seriousness of any particular breach. For example, in *Murray*, Murray was MFC's chief executive director. His service agreement contained an agreed damages clause requiring MFC to pay to Murray a year's gross salary if MFC wrongfully terminated his employment without providing 12 months' notice. MFC provided Murray with only 7.5 weeks' notice. Murray argued that MFC had wrongly terminated his employment such that he was entitled to the agreed damages amount. MFC challenged the clause as penal. Arden LJ held that MFC's obligation to pay for failing to give 12 months' notice was prorated for any notice given, and salary paid, within the requisite notice period.[66] Accordingly, properly interpreted, the clause required MFC to pay to Murray 44.5 weeks' gross salary.[67]

In some cases, Anglo-Australian courts choose not to *avoid* the rule; rather, they opt to *evade* it. English and Australian courts elide the rule in the same factual scenario, as follows.

An agreed compensatory damages clause in contract between A and B may be calculated based on A's anticipated 'bargain damages': that is, damages equivalent to the full value of the lost performance. In this thesis, an 'agreed bargain damages clause' refers to this type of clause.

---

[63] *Cenargo Ltd v Empresa Nacional Bazan de Construcciones Navales Militares SA* [2002] EWCA Civ 524, [2002] CLC 1151 [27]-[33]; *Pigram v A-G (NSW)* (1975) 132 CLR 216 (HCA) 221-22, 224-25, 227.

[64] (1993) 61 BLR 41 (PC) 54, 58-59.

[65] *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [75(iv)].

[66] *Murray v Leisureplay plc* [2005] EWCA Civ 963, [2005] IRLR 946 [75].

[67] Courts do not always take this approach, however: see *Arlesheim v Werner* [1958] SASR 136 (SASC) 141.

So, assume that a contract between A and B contains an agreed bargain damages clause. Assume further that the clause provides that B's obligation to pay the agreed bargain damages arises upon A terminating the contract for B's breach. The contract lists several breaches upon which A may terminate. Some are repudiatory; others are not. The *contractual* analysis is simple: upon *any* breach of contract by B, A obtains an express contractual right to terminate. The *general law* analysis, however, is more complicated. Two possibilities arise. If B's breach is repudiatory, A's expectation loss arising from termination flows from B's wrong; therefore, the breach gives A the right *also* to terminate at law, and claim bargain damages. Conversely, if B's breach is not repudiatory, A's expectation loss flows from her election to terminate, not B's wrong; thus, the breach does *not* give A the right to terminate at law, and B is only liable for the loss flowing directly from his non-repudiatory breach.[68] Assessed as the contract date, a court cannot know whether B's breach will be repudiatory, nor not. Accordingly, the Indiscriminate Cause Test should apply, as the same amount—the agreed bargain damages—will be payable upon the happening of either of two events which will result in varying harm.

In England, the Indiscriminate Cause Test does not apply in this scenario where A *in fact* relies upon B's breach of a condition. In *BNP Paribas v Wockhardt EU Operations (Swiss) AG*, Christopher Clarke J held that the fact B had breached a condition, and hence committed a repudiatory breach, ruled out applying the Indiscriminate Cause Test.[69] The problem with this analysis is that the Penalty Rule is applied as at the contract date, such that the precise type of breach will then be unknown. Accordingly, the nature of the breach is an *ex post* fact that a court properly cannot take into account when evaluating a clause to determine its validity. On orthodox English law principles, if an express contractual right to terminate based on a non-repudiatory breach *may* activate an agreed bargain damages clause, presumptively it will be penal, and that the breach activating the clause 'might, *as a matter of fact*, give rise to an entitlement to loss of bargain damages is not sufficient' to save it.[70]

In Australia, courts do not apply the Indiscriminate Cause Test in this scenario *at all*.[71] In fact, Australian courts deem that A's expectation loss arising from termination flows from B's breach in *all cases*, despite the fact that B's non-repudiatory breach may ground A's right to

[68] *Lombard North Central plc v Butterworth* [1987] QB 527 (CA) 542-43; *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 175, 176, 186, 191.

[69] [2009] EWHC 3116 (Com Ct) [45].

[70] JW Carter, *Carter's Breach of Contract* (LexisNexis Butterworths 2011) [13-09] (emphasis added).

[71] *Amev Finance v Artes Studios Thoroughbred Pty Ltd* (1989) 15 NSWLR 564 (NSWCA) 574-75.

terminate. It follows that, when determining whether an agreed bargain damages clause is penal, in every instance Australian courts compare the agreed bargain damages amount with A's anticipated bargain damages.[72]

Carter and Peden seek to justify the Australian position as follows: 'The Australian courts, by focusing on termination as the event on which damages are payable, have been able to avoid the "weakest link" approach.'[73] For two reasons, this reasoning is flawed. The first is that it completely disregards authority. The second is that one cannot divorce A's termination from B's breach upon which the termination is based. It that was possible, a court would be unable to review the clause, for the requisite default would not exist. In truth, Australian courts evade the Indiscriminate Cause Test by invoking a fiction that all the breaches leading to A's exercise of her right to terminate will be repudiatory. No subject can be rationally organised, or intelligibly applied, if it uses fictions.

To sum up, the Indiscriminate Cause Test displays a blanket nature. A clause that indiscriminately applies a single sum to diverse breaches is presumptively penal. On the surface, Anglo-Australian courts purportedly maintain the test. In fact, the courts only selectively apply the test. This approach leads to two problems. For one, it creates uncertainty, because it is hard to predict when a court will circumvent the rule, and when it will not. For another, it lacks intellectual integrity, for purporting to maintain a rule but then finding astute ways to avoid applying it constitutes subterfuge. Subterfuge undermines the law's legitimacy. Thus, Anglo-Australian courts must confront the issue directly. They should either jettison the test, or reinterpret it so that its true scope of application is made clear, and then demand that it be respected.

D The Impossibility Test

Under the Impossibility Test, the fact that 'precise pre-estimation is almost an impossibility' will not render a clause penal. Rather, the clause more likely will be upheld. For this test to apply, at the contract date the loss must be 'impossible to forecast' or 'impossible to foresee'; the nature of the contract must make it 'practically impossible to make an accurate pre-estimate of [loss]'.[74]

---

[72] *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 194, 197, 204, 210, 215; *Esanda Finance Corp Ltd v Plessnig* (1989) 166 CLR 131 (HCA) 147-48.

[73] JW Carter and Elisabeth Peden, 'A Good Faith Perspective on Liquidated Damages' (2007) 23 JCL 157, 175.

[74] *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd* [1915] AC 79 (HL) 88, 95, 104.

Two questions must be asked. As conventionally understood, does the test work? If not, is there another view upon which the test can function in an intellectually sustainable way?

I Functionality

Estimation may be impossible because the type of likely loss cannot be *determined*, or because although the type of likely loss is determinable, it cannot be *quantified*. Each is considered below.

An anticipated 'loss' may be indeterminable because the event that one is contemplating cannot be described as a loss without a fictitious use of language. If a 'loss' is indeterminable, and hence impossible to estimate, there can be no yardstick against which to compare the agreed compensatory damages amount. If no yardstick for assessment exists, any attempt to apply a comparative test of validity in these cases inevitably will yield an unintelligible answer. For example, Rossiter contends that, where the anticipated loss is 'impossible to prove, the courts are content to assume that the agreed sum is not disproportionate to that loss'.[75] The writer's statement is unintelligible not because it is wrong—the adjectives 'right' and 'wrong' are meaningless here—but rather because it is conceptually incoherent. In truth, if the 'loss' is indeterminable, any comparative test of validity is inoperable, and any agreed compensatory damages clause cannot be logically evaluated. Put differently, the Impossibility Test is by definition dysfunctional: it is meaningless to talk of assessing an estimate of loss against a 'loss' that is impossible to assess.

Conversely, the anticipated loss may be determinable, but unquantifiable. For instance, the loss may be unquantifiable because it is non-pecuniary in nature. On an orthodox loss-based view, this proposition is questionable for two reasons.

For one, when assessing damages for loss of amenity arising from personal injuries, courts are able to value intangibles such as the lost use of an eye, an ear, and so on. If so, surely courts can value all non-pecuniary losses potentially arising from a breach of contract.

For another, assume a defendant violates a claimant's right but the breach does not cause a calculable loss. The analysis will differ depending on whether the correct benchmark for assessment is anticipated *legally-recoverable* loss, or anticipated *actual* loss. If the former standard applies, then on a conventional loss-based approach the law can award only nominal damages. The basis for this conclusion is as follows: the situation where general compensatory damages are incalculable is analogous to that where a claimant shows a loss, but adduces no

---

[75] Chris Rossiter, *Penalties and Forfeiture* (Lawbook Company 1992) 34.

evidence on quantum. In that scenario, only nominal damages are awarded.[76] Accordingly, assessed at the contract date, if the anticipated loss is unquantifiable, the agreed compensatory damages amount must be tested against the nominal damages that a court would award. If so, then contrary to Lord Dunedin's statement in *Dunlop*, it cannot be more likely that the clause will be upheld.[77] If the latter standard applies, then no coherent yardstick for comparison exists, as the loss is per hypothesis immeasurable. As illustrated regarding cases where the 'loss' is indeterminable, this renders the Impossibility Test empty.

As traditionally interpreted, then, the Impossibility Test is problematic. If the test is concerned with loss, then the result is either that Lord Dunedin's judgment in *Dunlop* is wrong, or the test is dysfunctional, as one cannot intelligibly ask whether the agreed compensatory damages sum reflects the anticipated 'loss'.

II Another view

If one adopts a loss-based interpretation of the Impossibility Test, it is dysfunctional, as it cannot provide an intelligible answer to any comparative inquiry. In this context as in many others, the modern focus on loss is a loss of focus. If one adopts a rights-based approach, however, the Impossibility Test can be rendered functional. On this view, the test is best explained as addressing cases concerning substitution for the value of an infringed right. In such cases, loss is impossible to estimate because no anticipated loss exists. However, as Lord Dunedin explained, that is 'no obstacle to the sum stipulated' being upheld,[78] provided that it properly mirrors the anticipated value of A's infringed right. Further, given the difficulties in quantifying the value of an infringed right, courts are predisposed to upholding agreed substitutive damages clauses. This rights-based interpretation provides the best understanding of the Impossibility Test. Three 'central' cases that provided the foundation of the modern Penalty Rule[79] illustrate this.

First, in *Dunlop*, the case that Lord Dunedin cited as authority for the Impossibility Test was *Clydebank Engineering and Shipbuilding Co Ltd v Don Jose Ramos Yzquierdo y Castaneda*.[80] Clydebank contracted to build four warships for the Spanish Government. The contract provided that Clydebank was to deliver each vessel by a particular date. The agreement

---

[76] eg *Greer v Alstons Engineering Sales and Services Ltd* [2003] UKPC 46, (2003) 147 SJLB 783 [6]-[9].

[77] *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd* [1915] AC 79 (HL) 88.

[78] *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd* [1915] AC 79 (HL) 87.

[79] Chris Rossiter, *Penalties and Forfeiture* (Lawbook Company 1992) 28.

[80] [1905] AC 6 (HL).

contained an agreed damages clause stipulating that if Clydebank delivered a ship late, it would be required to pay a 'penalty' at the rate of £500 per week for the period of its delay. Clydebank delivered each warship a number of months late. The Spanish Government claimed £500 for each week Clydebank had delayed in delivering each vessel. In response, Clydebank alleged the clause was correctly labelled.

The House of Lords held that the clause misdescribed the obligation to pay as a 'penalty'. In fact, it was a liquidated damages clause.

Lord Halsbury held that, as at the contract date, it was impossible to pre-assess the loss that might flow from any breach. His Lordship noted that, given warships are not profit-generating, the Spanish Government would suffer no loss from late delivery.[81] This must be correct: no anticipated loss existed because the Government would never be factually worse off as a result of Clydebank's breach. It seemingly follows that the anticipated 'loss' was impossible to estimate because it did not in fact exist.

Compare Lord Robertson's judgment. It is critical. His Lordship held that the relevant clause was enforceable if it constituted 'a genuine pre-estimate of the creditor's probable or possible interest in the due performance of the principal obligation', adding that 'Unless injury to a State is as [a] matter of law inexpressible in money, [the Spanish Government] was or might be … deeply injured by delay.'[82] In modern terms, his Lordship acknowledged that the relevant comparator was not the Spanish Government's anticipated loss, but rather the anticipated value of its infringed performance right. Lord Robertson's focus on the infringed performance right is clear by his reference to the Spanish Government's *injuria*, being the wrong of the rights violation, and not its *damnum*, being the consequential loss arising from that infringement, of which none would arise. This interpretation has support. Edelman once analysed *Clydebank* as a case where the remedy granted was 'explicable only as the value of the right', and not one concerning financial loss.[83]

If this is correct, then *Clydebank* is best understood as follows. The House of Lords upheld the agreed substitutive damages clause because the agreed damages amount was not disproportionate to the anticipated value of the infringed performance right. Given at the time there existed no readily available market for the sale and purchase of warships, the value of the

---

[81] *Clydebank Engineering and Shipbuilding Co Ltd v Don Jose Ramos Yzquierdo y Castaneda* [1905] AC 6 (HL) 12.

[82] *Clydebank Engineering and Shipbuilding Co Ltd v Don Jose Ramos Yzquierdo y Castaneda* [1905] AC 6 (HL) 19-20.

[83] James Edelman, 'Gain-Based Damages and Compensation' in Andrew Burrows and Alan Rodger (eds), *Mapping the Law: Essays in Memory of Peter Birks* (OUP 2006) 156.

right was quantified adopting a reasonable, hypothetical bargain approach. So, the question posed, at the contract date, was this: in a future hypothetical bargain between a reasonable, willing government and a reasonable, willing warship builder, would the government release its right to timely performance for a rate of somewhere generally in the vicinity of £500 per week for the period of delay? The answer, seemingly, was 'yes'.

Second, in *Comr of Public Works v Hills*, delivering the Privy Council's advice Lord Dunedin held:

> The *general principle to be deduced from* [*Clydebank*] seems to be … that the criterion of whether a sum … is truly liquidated damages … is to be found in whether the sum stipulated for can or cannot be regarded as a genuine pre-estimate of the creditor's probable or possible interest in the due performance of the principal obligation.[84]

Three points are noteworthy. To start, the principal architect of the modern Penalty Rule produced this passage. Next, according to Lord Dunedin, the relevant question when applying the Impossibility Test is whether the agreed damages amount is an appropriate pre-assessment of the value of A's infringed performance right. Lastly, the Australian High Court has applied this passage,[85] and Anglo-Australian appellate courts have continually cited it with approval.[86]

Finally, one must return to *Dunlop*. In that case, the House of Lords relevantly held that the agreed damages clause passed the Impossibility Test. On this issue, Lord Atkinson's judgment is seminal, for two reasons. On the one hand, his Lordship phrased the Estimation Test in slightly different terms to Lord Dunedin—a liquidated damages clause represents a genuine pre-estimate of A's 'probable or possible interest in the due performance of the contract'.[87] On the other hand, Lord Atkinson held that Dunlop's object in entering the relevant contract was to prevent undercutting, which would disorganise its trading system. His Lordship further held that Dunlop had an interest in preventing undercutting, and that it was 'impossible to say that that interest was incommensurate with the sum agreed to be paid'.[88] According to the Australian High Court in *Andrews v Australia and New Zealand Banking Group Ltd*, it followed that in *Dunlop*, 'the critical issue … was whether the sum agreed was

---

84 [1906] AC 368 (PC) 375-76 (emphasis added).

85 *Lamson Store Service Co Ltd v Russell Wilkins and Sons Ltd* (1906) 4 CLR 672 (HCA) 681-82, 686, 688-89.

86 *Hamilton v Lethbridge* (1912) 14 CLR 236 (HCA) 264; *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [77].

87 *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd* [1915] AC 79 (HL) 96.

88 *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd* [1915] AC 79 (HL) 92.

commensurate with the interest protected by the bargain'.[89]

As the Australian High Court has previously pointed out, the test of validity is not concerned with the *interests* a contract protects.[90] It is, however, concerned with *rights*. It is well known that interests and rights are not synonymous.[91] However, many rights 'do take the reason for their existence from a particular interest in the world to which they relate'.[92] In *Dunlop*, Dunlop had an interest in resale price maintenance. That interest formed the subject matter of Dunlop's performance right, and provided the reason for why its performance right existed. Properly viewed, then, the 'critical issue' in *Dunlop* was whether the agreed substitutive damages clause properly reflected the anticipated value of the infringed performance right, as measured by a hypothetical, reasonable bargain for the release of the covenant. It did.

In conclusion, it has been said that the rights-based approach to remedies 'offers explanations of legal rules that are otherwise difficult to justify'.[93] The Impossibility Test is a primary example.

E The Extravagance Test

I The primary or sole test

The Extravagance Test provides that a sum is penal if it is 'extravagant and unconscionable in amount' compared to 'the greatest loss that could … have followed from the breach'.[94] In recent years, Anglo-Australian case law has shifted such that effectively the Extravagance Test is now the sole test of validity. The Australian authorities are straightforward. Though the English authorities are more complex, they are best interpreted in precisely the same way.

At the time the Australian High Court decided *AMEV-UDC Finance Ltd v Austin*,[95] Australian intermediate appellate court authorities emphasised that the Estimation Test was primary, and that it required a substantially accurate estimation of the anticipated loss. In *AMEV,* Mason and Wilson JJ held that courts 'should return to the *Clydebank* and *Dunlop*

---

89 [2012] HCA 30, (2012) 247 CLR 205 [75].

90 *Ringrow Pty Ltd v BP Australia Pty Ltd* [2005] HCA 71, (2005) 224 CLR 656 [27].

91 JE Penner, 'The Analysis of Rights' (1997) 10 Ratio Juris 300, 306-307.

92 Robert Stevens, 'Rights and Other Things' in Donal Nolan and Andrew Robertson (eds), *Rights and Private Law* (Hart Publishing 2012) 133.

93 Donal Nolan and Andrew Robertson, 'Rights and Private Law' in Donal Nolan and Andrew Robertson (eds), *Rights and Private Law* (Hart Publishing 2012) 20.

94 *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd* [1915] AC 79 (HL) 87.

95 (1986) 162 CLR 170 (HCA).

concept', and hold an agreed damages amount penal only if it was 'out of all proportion to damage likely to be suffered as a result of breach'.[96] Two points are important. For one, their Honours cited the Law Commission, which had argued that a test of validity requiring a substantially accurate estimate was unrealistic.[97] For another, at the time the English cases had already used the language of 'out of all proportion'[98] and 'disproportionality'[99] when discussing the test of validity.

In *Ringrow Pty Ltd v BP Australia Pty Ltd*,[100] the High Court completed the (re)interpretive process that *AMEV* had begun. The High Court held that the *Dunlop* judgment represented the law in Australia. The Court then emphasised a 'reformulated'[101] Extravagance Test: namely, an agreed damages sum is penal only if it is 'out of all proportion' with the anticipated likely loss.[102] The critical issue, then, is the degree of disproportion between the agreed damages sum (or value) and the amount of the likely future loss. Importantly, in *Ringrow*, the High Court made little mention of the Estimation Test. Also, it stated that a penalty clause is one creating 'advantages *significantly greater* than the advantages flowing from a genuine pre-estimate of damage'.[103]

Since *Ringrow*, Australian intermediate appellate courts have applied the reformulated Extravagance Test. The most important case to do so is *Zachariadis*.[104] In that case, the Victorian Court of Appeal considered whether there is a sole applicable test in deciding whether an agreed damages clause is penal. After an extensive survey of the Australian authorities, the Court concluded that it was the 'out of all proportion' test.[105] The test is generally applicable, but it is applied contextually, as the 'factual circumstances in which the contract is made will affect the application of that test'.[106]

---

[96] *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 190.

[97] Law Commission, *Penalty Clauses and Forfeiture of Monies Paid* (Law Com WP No 61, 1975) [33].

[98] *Stockloser v Johnson* [1954] 1 QB 476 (CA) 490.

[99] *Gilbert-Ash (Northern) Ltd v Modern Engineering (Bristol) Ltd* [1974] AC 689 (HL) 711.

[100] [2005] HCA 71, (2005) 224 CLR 656.

[101] JW Carter and Elisabeth Peden, 'A Good Faith Perspective on Liquidated Damages' (2007) 23 JCL 157, 167.

[102] *Ringrow Pty Ltd v BP Australia Pty Ltd* [2005] HCA 71, (2005) 224 CLR 656 [32].

[103] *Ringrow Pty Ltd v BP Australia Pty Ltd* [2005] HCA 71, (2005) 224 CLR 656 [27] (emphasis added). See similarly *Scandinavian Trading Tanker Co AB v Flota Petrolera Ecuatoriana (The Scaptrade)* [1983] 2 AC 694 (HL) 702.

[104] See also *Riggall v Thompson* [2010] QCA 144 [31], [34]; *Birdanco Nominees Pty Ltd v Money* [2012] VSCA 64, (2012) 219 IR 276 [87]; *Spiers Earthworks Pty Ltd v Landtec Projects Corporation Pty Ltd (No 2)* [2012] WASCA 53, (2012) 287 ALR 360 [21].

[105] *Zachariadis v Allforks Australia Pty Ltd* [2009] VSCA 258, (2009) 26 VR 47 [141]-[143].

[106] *Zachariadis v Allforks Australia Pty Ltd* [2009] VSCA 258, (2009) 26 VR 47 [143].

In short, in Australia, the Extravagance Test is the sole test of validity.

Turning to English law, recent appellate authority has emphasised the primacy of the Extravagance Test. The first important case was *Murray*. The English Court of Appeal unanimously held that the relevant clause provided for liquidated damages, but it famously split in its reasoning. On the one hand, Arden LJ favoured a test of validity that compared the agreed damages amount with the general damages a court would award, and then required A to justify any discrepancy between the two amounts.[107] On the other hand, Buxton LJ disagreed with Arden LJ's approach. His Lordship emphasised that the Extravagance Test is the 'principal test', such that a clause is only penal if the stipulated sum fails that test; and that a court will not impugn an agreed damages sum merely because it is 'generous'.[108] Clarke LJ held that he preferred Buxon LJ's 'broader approach'.[109] According to Peel, the 'broader approach' enunciated in *Murray* 'places most, if not all, of the emphasis' on the Extravagance Test.[110]

Later, in *Euro London Appointments Ltd v Claessens International Ltd*, the English Court of described the Extravagance Test as 'the classic statement' of the relevant test of validity, and assessed the relevant terms against a standard of extravagance.[111]

The latest word on the issue is *Makdessi*. After considering the 'broader approach', the English Court of Appeal held that it did not interpret *Murray* to mean that 'the only question for consideration' is whether the clause passed the Extravagance Test.[112] In form, the Court appeared to reject the Extravagance Test as the sole test. In substance, it held precisely the opposite. The reasoning to this conclusion proceeds in three steps.

First, in the Court of Appeal's view, to hold that the Extravagance Test was the sole test would be to 'ignore' the remainder of Lord Dunedin's speech in *Dunlop* and Lord Woolf's opinion in *Philips*.[113] That is not quite right. As to the former, this chapter has illustrated that, in truth, the other potential tests of validity are inadequate, obsolete, or are best reinterpreted. As to the latter, it is unclear how holding that the Extravagance Test is the sole test would disregard *Philips*. This only follows if one reads the judgment as preferring the Estimation Test to the Extravagance Test. However, this is impossible, as the Privy Council (mistakenly)

---

[107] *Murray v Leisureplay plc* [2005] EWCA Civ 963, [2005] IRLR 946 [54].

[108] *Murray v Leisureplay plc* [2005] EWCA Civ 963, [2005] IRLR 946 [114]-[115].

[109] *Murray v Leisureplay plc* [2005] EWCA Civ 963, [2005] IRLR 946 [105].

[110] Edwin Peel, *The Law of Contract* (13th edn, Sweet and Maxwell 2011) [20-139].

[111] [2006] EWCA Civ 385, [2006] 2 Lloyd's Rep 436 [15], [37]-[39].

[112] *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [98].

[113] *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [98].

viewed the two tests as coterminous.[114] Rather, to hold that the Extravagance Test is *not* the sole test would be to 'ignore' the majority view of the Court of Appeal in *Murray*, a move that the Court of Appeal in *Makdessi* would not be entitled to make.[115]

Second, despite the Court of Appeal's express statement, in other passages it asserts that the Extravagance Test is the sole, or primary, test. For example, the Court concluded that the authorities showed courts adopting 'the broader test' of asking whether the clause was 'extravagant and unconscionable'; and that this constituted a 'reversion to the foundation' of the Penalty Rule, which required a court to ask whether the agreed damages clause was 'unconscionable and extravagant'.[116]

Finally, on the surface, the Court of Appeal applied the Estimation Test. However, a close examination of the way in which the court conducted its analysis reveals the truth: the Court actually applied the Extravagance Test, and used the phrase 'genuine pre-estimate' in a conclusory way, applying it as an *ex post facto* label once the analysis was complete. For instance, in purportedly applying the Estimation Test, the Court: applied the Extravagance Test in its precise terms; tested the relevant clause against standards of 'extravagance' and 'total disproportionality'; and applied the Extravagance Test in its reformulated guise, holding that the relevant clause would likely produce a result that was 'totally out of proportion'.[117] Further, the Court held that where a clause was extravagant based on this process of reasoning, it *then* properly was not a 'genuine pre-estimate'.[118] Accordingly, the court used the 'genuine pre-estimate' phrase as no more than a conclusory label with no normative import.

In summary, the contemporary Anglo-Australian authorities focus on the degree of disproportion between the agreed compensatory damages amount and the anticipated loss. Properly interpreted, courts no longer analyse clauses in terms of pre-estimates, genuine or otherwise. Consequently, as best understood, the Penalty Rule contains one test of validity: the Extravagance Test.

II Interpretation

---

[114] Text to n 32.

[115] *Young v Bristol Aeroplane Co Ltd* [1944] KB 718 (CA) 729, affd [1946] AC 163 (HL).

[116] *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [104]. See also [120].

[117] *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [107], [110], [112].

[118] *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [105], [117].

In its original form, the Extravagance Test provides that an amount is penal if it is extravagant *and unconscionable* in amount. There exists some ambiguity as to what 'unconscionability' means. The authorities and literature reveal at least three interpretations. The term may encapsulate the principle that a party cannot unconscionably enforce a strict legal right. The phrase may be an historical reference, signifying an inequality of bargaining power. Or it may be a synonym for the expression 'extravagant'. Each interpretation is considered in sequence.

(1) Unconscionable enforcement of a strict legal right

The 'unconscionable enforcement' principle holds that a party cannot unconscionably enforce a strict legal right. This is said generally to provide a 'fundamental principle' grounding equitable intervention.[119] Courts and commentators have asserted that the Penalty Rule exemplifies equity's broad concern to prevent the unconscientious exercise of legal rights. For example, in *Scandinavian Trading Tanker Co AB v Flota Petrolera Ecuatoriana (The Scaptrade)*, Goff LJ held that the Penalty Rule ensures that a claimant cannot 'use [her] legal rights to take advantage of another's misfortune …'.[120] Accordingly, so the argument runs, the term 'unconscionable' in the Extravagance Test encapsulates the unconscionable enforcement principle.

This interpretation is flawed. It is presented as an explanation of how the Extravagance Test operates, but it does not—and cannot—explain anything. Three reasons follow.

First, the content of the epithet 'unconscionable' in the unconscionable enforcement principle is indeterminate. As Worthington notes, the term 'unconscionable' is 'hardly helpful without a lot more explanation'.[121] Indeed, in *Royal Brunei Airlines Sdn Bhd v Tan*, Lord Nicholls noted that the word 'unconscionable' is 'not a word in everyday use by non-lawyers' and that it was 'better avoided', given the difficulty in identifying precisely what that term meant.[122] Embedding the uncertain concept in a particular context—that of the enforcement of strict legal rights—does not resolve its inherent definitional issues. Faced with no fixed meaning of the term, some writers have resorted to the safe haven of complete abstraction. So, Lanyon

---

[119] *Legione v Hateley* (1983) 152 CLR 406 (HCA) 444. See similarly *Bridge v Campbell Discount Co Ltd* [1962] AC 600 (HL) 629.

[120] [1983] QB 529 (CA) 538, affd [1983] 2 AC 694 (HL). See also Chris Rossiter, *Penalties and Forfeiture* (Lawbook Company 1992) 142.

[121] Sarah Worthington, *Equity* (2nd edn, OUP 2006) 199.

[122] [1995] 2 AC 378 (HL) 392.

argues that the Extravagance Test refers to an unconscionability concept that operates 'at large'.[123]

The issue is obvious. To say that unconscionability is 'at large' renders the term completely unspecific. Unless 'unconscionability' is more specifically defined, it is simply a 'category of meaningless reference',[124] because it could mean anything. Worse still, it has the potential to mask an intuitive evaluation. Perhaps for this reason, the Australian High Court has recently recognised that *any* unconscionability concept does not, and cannot, operate wholly 'at large'.[125] The problem is this. Some commentators insist that the concept of an 'unconscionable' enforcement of a legal right has substantive meaning. We now know what that meaning is *not*. However, it remains totally obscure as to what that substantive meaning *is*.

Second, the notion of 'unconscionability' in the unconscionable enforcement principle is conclusory, not explanatory. In *Garcia v National Australia Bank Ltd*, Gaudron, McHugh, Gummow and Hayne JJ held that 'the statement that enforcement of the transaction would be "unconscionable" is to characterise the result rather than to identify the reasoning that leads to the application of that description'.[126] This interpretation of the Extravagance Test is open to the same criticism. For example, in *AMEV*, Mason and Wilson JJ held that A will act unconscionably if she seeks to enforce a disproportionate term.[127] If A has acted unconscionably, she has done so *ex post*, in attempting to enforce a clause of a transaction that was not entered into unconscionably. But this merely reflects the antecedent determination that the clause must not be enforced due to its disproportionality. It is therefore the disproportionality that renders the attempted enforcement 'unconscionable'. The reason for the clause being impugned cannot be found in any notion of 'unconscionability'. Accordingly, unconscionability *ex post* has no explanatory weight.

Finally, in *Union Eagle Ltd v Golden Achievement Ltd,* Lord Hoffmann held that two practical business considerations justified rejecting the unconscionable enforcement principle.[128] An 'undefined discretion' to refuse to enforce an express term in a commercial contract on unconscionability grounds would create uncertainty. Further, even if courts rarely

---

123 Elisabeth Lanyon, 'Equity and the Doctrine of Penalties' (1996) 9 JCL 234, 246.

124 *Australian Broadcasting Corp v Lenah Game Meats Pty Ltd* [2001] HCA 63, (2001) 208 CLR 199 [99] (footnote omitted).

125 *Tanwar Enterprises Pty Ltd v Cauchi* [2003] HCA 57, (2003) 217 CLR 315 [20], [86].

126 [1998] HCA 48, (1998) 194 CLR 395 [34].

127 *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 193.

128 [1997] AC 514 (PC) 519.

exercised that discretion, its *mere existence* would enable one party to use litigation as a commercial 'negotiating tactic'.

In short, Anglo-Australian private law should heed Lord Nicholls's warning. This interpretation is best avoided.

(2) Inequality of bargaining power

The expression 'unconscionable' in the Extravagance test may be an historical reference. As Chapter 2 explained, historically on occasion courts suggested that they were prepared to impugn a clause if it was 'oppressive' in nature, or if there existed a 'gross' inequality between the parties.[129] Accordingly, Harris, Campbell and Halson state that 'the historical explanation' for the Penalty Rule is that it sought to redress an inequality of bargaining power.[130] On this view, then, the term 'unconscionable' in the Extravagance Test refers to an inequality of bargaining power.

Several judicial statements suggest that the rule is concerned with redressing bargaining power inequalities. For example, in *Schroeder Music Publishing Co Ltd v Macaulay*, Lord Diplock held that the Penalty Rule applies a principle of protecting 'those whose bargaining power is weak' from being forced to enter into an unconscionable bargain.[131] Many academics take the same view. For instance, Harris argues that the rule seeks to 'protect a weaker party against the … use of a superior bargaining strength'.[132]

Though this interpretation appears transparent, it fails for at least three reasons.

To start, this interpretation lacks fit. The expression 'unconscionability' cannot refer to an abuse of bargaining power because such conduct is not necessary or sufficient to pass the Extravagance Test. As Keene LJ has stated, 'the concept of a penalty clause is not confined to situations where one party had a dominant bargaining power over the other … the situation where one party is dominant is not exhaustive of those contracts where a penalty may be identified'.[133]

---

[129] Text to n 41 and n 43 in ch 2.

[130] Donald Harris, David Campbell and Roger Halson, *Remedies in Contract and Tort* (2nd edn, Butterworths 2002) 138.

[131] [1974] 1 WLR 1308 (HL) 1315. See also *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 194.

[132] Donald Harris, 'Incentives to Perform, or Break Contracts' (1992) 45 CLP 29, 39.

[133] *Jeancharm Ltd v Barnet Football Club Ltd* [2003] EWCA Civ 58 [20]-[21].

Many cases exist where the parties were of roughly equal bargaining power, but the clause was nonetheless held penal. For instance, in *Makdessi*, Makdessi was a 'key figure' in the Middle East advertising and marketing world; Cavendish was a holding company within the world's largest market communications services group; and experienced lawyers for both sides had extensively negotiated the clause.[134] Although a level playing field existed, the clause was invalidated.

Conversely, other cases exist where the parties were *not* of an equal bargaining strength, yet the agreed damages clause was upheld. For example, in *Philips*, the relevant contract was mostly on the Government's standard terms; the Government supplied the agreed damages amount in the tender forms; and Philips did not have a chance to negotiate the agreed damages clause.[135] This suggests the Government took advantage of a dominant bargaining position, but the Privy Council nevertheless held the clause valid.

Next, although this interpretation is seemingly transparent, many counterexamples exist. For instance, in *Imperial Tobacco Co (of Great Britain and Ireland) Ltd v Parslay*, Lord Wright MR held that the use of the term 'unconscionability' in the Extravagance Test does not refer 'to the fact that the parties were on an unequal footing. It does not bring in at all the idea of an unconscionable bargain.'[136] Similarly, in *Citicorp Australia Ltd v Hendry*, Kirby P held that the test of validity paid no 'regard to the relative bargaining positions of the parties'.[137] Simply put, the test of validity is 'not concerned with the general position of the parties vis-à-vis each other'.[138]

Lastly, the dicta and comments seemingly supporting this interpretation are explicable. Courts and commentators apparently have been led astray by the fact that many of the leading cases involved facts displaying institutional inequality.[139] For example, several cases concerned hire purchase contracts, where a finance company usually will be in a dominant position.[140] However, even if it was *descriptively* accurate to say that inequality existed in those cases, it was *normatively* insignificant. Its existence was merely coincidental: if no inequality had existed, the outcome would have been the same.

---

134 *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [2], [4], [5].

135 *Philips Hong Kong Ltd v A-G of Hong Kong* (1993) 61 BLR 41 (PC) 50-51.

136 [1936] 2 All ER 515 (CA) 521. See also 525-26.

137 (1985) 4 NSWLR 1 (NSWCA) 23.

138 *Jeancharm Ltd v Barnet Football Club Ltd* [2003] EWCA Civ 58 [26].

139 eg Hugh Beale, *Remedies for Breach of Contract* (Sweet and Maxwell 1980) 59.

140 eg *Bridge v Campbell Discount Co Ltd* [1962] AC 600 (HL).

(3) Extravagant

The final potential interpretation is that 'unconscionable' as it appears in the Extravagance Test is merely a synonym for the expression 'extravagant'. Much support for this view exists, in England and Australia.

As to the former, in *Dunlop*, Lord Dunedin's use of the phrase 'unconscionable *in amount'* suggests that the entire phrase should be read as a synonym for extravagant; whilst Lord Atkinson referred to a penalty clause as being 'unreasonable, unconscionable, or extravagant', suggesting that the terms were interchangeable.[141] Later, in *Imperial Tobacco*, Lord Wright MR explained that the reference to 'unconscionability' was 'merely a synonym for something which is extravagant and exorbitant'; whilst in Romer LJ's view, the 'only question' to be determined was whether the stipulated sum was 'obviously larger' than the anticipated loss.[142] Recently, in *Makdessi*, the Court of Appeal held that 'the term "unconscionable" would perhaps more appropriately be used for a clause which provides for extravagant payment …'.[143]

As to the latter, in *Ringrow*, the High Court held that the phrase 'out of all proportion' captured the meaning of phrase 'extravagant and unconscionable in amount', suggesting that the terms were synonymic; and it spoke of the need for the discrepancy *in value* to be 'extravagant and unconscionable'.[144]

On this view, the term 'unconscionable' has no independent content. If that is correct, the Penalty Rule's operation in cases of derisory agreed damages clauses is justified. Accordingly, the Extravagance Test displays a tighter fit than the Estimation Test. To explain, where an agreed compensatory damages sum is disproportionately small in relation to the anticipated loss, it is not a 'genuine pre-estimate', and so it fails the Estimation Test. However, the sum is *not* extravagant, and so it will pass the Extravagance Test. Why? This is because the very notion of 'extravagance' moves in one direction only. Where the agreed damages are

---

[141] *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd* [1915] AC 79 (HL) 87 (emphasis added), 97.

[142] *Imperial Tobacco Co (of Great Britain and Ireland) Ltd v Parslay* [1936] 2 All ER 515 (CA) 521, 526.

[143] *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [125].

[144] *Ringrow Pty Ltd v BP Australia Pty Ltd* [2005] HCA 71, (2005) 224 CLR 656 [21], [32]. See also *Calcorp (Australia) Pty Ltd v 271 Collins Pty Ltd* [2010] VSCA 259, (2010) 29 VR 462 [34]-[35].

'underestimated', any question of the amount being extravagant 'does not arise'.[145] It follows that *Cellulose* and its progeny were correctly decided.

Note that a derisory agreed damages clause is accurately described as being 'disproportionate'. That is, a derisory agreed compensatory damages sum may be 'out of proportion' to the anticipated loss because it is comparatively too low. Thus, it is disproportionately *small*. As such, on this view a derisory agreed damages clause should be invalid. But that is not so. Accordingly, it is better to use a term that necessarily contains within it a limit to disproportionately *large* sums. It follows that the language of 'extravagance' is preferable to that of 'disproportionality'.

To sum up, the expression 'unconscionable' in the Extravagance Test is best interpreted as being a synonym for 'extravagant'.

2 Problematic issues

The case law regarding the Penalty Rule test of validity raises three seemingly difficult issues. Is the test of validity subject to a 'commercial justification' qualifier? When applying a comparative test to an agreed compensatory damages clause, is the proper benchmark for assessment anticipated actual, or legally-recoverable, loss? What is the correct assessment date, and how does a court properly apply a test of validity operating as at that date? An answer to each question follows.

A The 'commercial justification' qualifier

In *Makdessi*, the English Court of Appeal held that recent authority heralded a 'new approach' to the test of validity.[146] The new approach involves imposing a qualification or 'cut back' on the traditional test of validity. This section begins with a distinctly English story of that qualifier. It does not have a happy ending. The section then provides a critical review of the story.

I The test

145 *Widnes Foundry (1925) Ltd v Cellulose Acetate Silk Co Ltd* [1931] 2 KB 393 (CA) 415, affd [1933] AC 20 (HL). See similarly *Multiplex Constructions Pty Ltd v Abgarus Pty Ltd* (1992) 33 NSWLR 504 (NSWSC) 519.

146 *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [84].

The story begins with *Lordsvale Finance plc v Bank of Zambia*.[147] A borrower entered into a loan agreement with a lender. A clause in the agreement provided that if the borrower defaulted, it was to pay interest on the outstanding amount at a rate of one per cent in addition to the interest rate ordinarily payable. The rate increase was unexplained. The borrower defaulted. The lender sought to recover its default interest. The borrower claimed the clause was penal.

In principle, Colman J held that even if an agreed damages clause fails the Estimation Test, a court should nevertheless uphold it if the discrepancy between the agreed damages amount and anticipated loss can 'in the circumstances be explained as commercially justifiable'.[148] In this thesis, the 'Commercial Justification Test' refers to this test. On the facts, Colman J held that the clause was properly a liquidated damages clause. The judge reasoned that, although the default interest rate increase was not 'in any real sense a genuine pre-estimate of loss', the increase was commercially justified, as a borrower in default poses an increased credit risk.[149]

English appellate courts have continually cited with approval the Commercial Justification Test. For example, in *Murray*, Arden LJ held that if B demonstrates that the agreed damages amount 'does not constitute a genuine pre-estimate of loss', then A may save the clause by illustrating that there is 'some other reason which justifies the discrepancy' between the two amounts.[150] Recently, in *Makdessi*, the Court of Appeal powerfully confirmed that the Commercial Justification Test forms part of the new approach to the test of validity.[151]

Gullifer supports the commercial justification qualifier. The writer argues that courts uphold agreed damages clauses if 'there is a commercially justifiable reason for them', and that they are justified in doing so.[152]

In England, then, a proper 'commercial justification' can save a prima facie penalty clause. There is no hint of this in Australian law at the appellate level.

II Evaluation

---

[147] [1996] QB 752 (QB).

[148] *Lordsvale Finance plc v Bank of Zambia* [1996] QB 752 (QB) 764.

[149] *Lordsvale Finance plc v Bank of Zambia* [1996] QB 752 (QB) 763, 767.

[150] *Murray v Leisureplay plc* [2005] EWCA Civ 963, [2005] IRLR 946 [54(v)]. See also [38], [42] (Arden LJ), [106(v)] (Clarke LJ).

[151] *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [88], [104], [117], [125].

[152] Louise Gullifer, 'Agreed Remedies' in Andrew Burrows and Edwin Peel (eds), *Commercial Remedies: Current Issues and Problems* (OUP 2003) 198.

The Commercial Justification Test is desperately flawed. For at least four reasons, the test cannot be accepted.

First, there is no interpretive warrant for a cut back of the kind the Commercial Justification Test provides. The test does not naturally evolve from the 'central' cases, such as *Dunlop*, which do not even intimate that a prima facie invalid clause might be saved. It does not constitute an exercise in interpreting the present test of validity; it constitutes inventing a new one.

Second, if one applies commercial justification reasoning to many of the decided cases, it will generate outcomes at odds with the actual results. In short, it fits poorly with the current law.

Take, for instance, *Jeancharm Ltd v Barnet Football Club Ltd*. Jeancharm was a clothing manufacturer. It contracted to supply football kits to Barnet, a football club. One clause provided that if Jeancharm delivered any items late, it would pay a 'penalty' of 20 pence per day for each garment not supplied. Another provision stipulated that if Barnet were late in paying its invoices, it would pay a 'late payment penalty' of interest on the outstanding sum accruing at 260 per cent per annum. Barnet defaulted. Jeancharm sought to invoke the default interest clause. Jeancharm argued that the default interest clause was balanced by the late delivery clause. The trial judge had upheld the default interest rate clause, holding that it was a 'quid pro quo for the late delivery clause'.[153] Surely, if anything provides a 'commercial justification' for a prima facie penalty clause, it is the existence of a parallel penalty clause. However, the English Court of Appeal unanimously held that the default interest clause was penal. In particular, Peter Gibson LJ held that 'the arguably penal terms of the late delivery clause [do not] justify the penal interest clause'.[154] If the Commercial Justification Test is right, then the decision was wrong.

Now consider Arden LJ's judgment in *Murray*. Her Ladyship held that there existed a discrepancy between the agreed damages amount and any likely general damages award, such that the relevant clause failed the Estimation Test. However, Arden LJ held that several factors commercially justified the discrepancy, thus saving the clause.[155] Three factors either jointly or severally provided sufficient commercial justification: the agreed damages accorded with market expectations; the contract included 'some fierce restrictions on competition'; and

---

[153] *Jeancharm Ltd v Barnet Football Club Ltd* [2003] EWCA Civ 58 [26].

[154] *Jeancharm Ltd v Barnet Football Club Ltd* [2003] EWCA Civ 58 [26].

[155] *Murray v Leisureplay plc* [2005] EWCA Civ 963, [2005] IRLR 946 [54], [69]-[76].

MFC obtained significant advantages from the clause—specifically, it allowed MFC to 'effect a clean break' and know its measure of financial exposure.[156]

In truth, Arden LJ's 'commercial justification' factors were empty. As to the first factor, a court will take into account market expectations when determining the anticipated loss under the comparative test of validity. Only such an approach can satisfy the requirement that the test of validity is applied contextually.[157] Market expectations in the abstract cannot justify anything. As explained below, the second factor concerning the stringent restrictive covenants is properly relevant to assessing A's anticipated mitigation opportunities, not any commercial justification. The final factor is fundamentally flawed. All agreed damages clauses—whether liquidated damages clauses *or penalty clauses*—will effect a clean break, and allow B to know the measure of his exposure. If this final factor suffices as a commercial justification, then *all* agreed damages clauses would be valid. That is not the law. Accordingly, then, applying her Ladyship's principles correctly, the relevant clause failed the Estimation Test *and* the Commercial Justification Test (as no proper commercial justification existed). The clause was therefore invalid. The majority held otherwise.

Third, the Commercial Justification Test lacks explanatory power. In *Makdessi*, the Court of Appeal put forward several cases where a commercial justification allegedly supported the decision.[158] In fact, the outcomes in those cases are better explained on other bases. Two examples suffice to demonstrate the point.

One case is *Lordsvale*. However, properly analysed, the relevant clause *passed* the test of validity, even if one applies the Estimation Test. In *Jeancharm*, Jacob J interpreted Colman J in *Lordsvale* as having held that the one per cent uplift was 'a genuine pre-estimate of loss on the basis that it indicated that the borrower was a risky borrower'.[159] Jacob J suggested that in cases involving a small default interest rate increase, the additional interest is justified as compensating for anticipated loss. In strikingly similar circumstances, Australian courts have adopted that precise reasoning.[160] So, the increased credit risk meant that the lender's anticipated loss was higher, such that the new rate properly reflected that potential loss,

---

[156] *Murray v Leisureplay plc* [2005] EWCA Civ 963, [2005] IRLR 946 [71], [74].

[157] *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd* [1915] AC 79 (HL) 87.

[158] *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [121].

[159] [2003] EWCA Civ 58 [16].

[160] *David Securities Pty Ltd v Commonwealth Bank of Australia* (1990) 23 FCR 1 (FCAFC) 31, revd (1992) 175 CLR 353 (HCA).

assessed as at the contract date. In short, the result is justified by orthodoxly applying the test of validity.

Another case is *Euro London*. It is said that the commercial justification for the relevant clause was that it allocated credit risk to the client rather than to the agency. But if the allocation of a *credit* risk provides a sufficient commercial justification, then so too must the allocation of a risk *generally*. If so, all agreed damages clauses are valid, for in *every case* A may argue that the clause represents a negotiated allocation of risk between the parties. The exception, then, will have swallowed up the rule.[161] One justifies the outcome at the expense of the rule's very existence. The English Court of Appeal's actual reasoning in the case more soundly explains the result: the relevant clause involved a condition precedent to a contractual right, and hence it was not an agreed damages clause at all.[162]

Finally, the Commercial Justification Test is illegitimate, as it allows for policy-based reasoning. It has been held that the commercial justification may be an 'economic' one.[163] For example, A might save a prima facie penalty clause by arguing that it will result in efficiency gains. Further, it has been held that public interest concerns count. For instance, A may rescue a prima facie penalty clause by establishing that a 'paramount public interest' consideration requires that the clause be upheld.[164] As Chapter 1 explained, to apply such a test, judges would be required to make decisions based on policy concerns that are non-justiciable. Judges lack the technical competence to assess the merits of economic efficiency arguments. Also, courts cannot be expected to weigh incommensurable policy considerations to determine whether a particular public interest is 'paramount'.

In conclusion, this English story is irreparably flawed at its base. It is a false step that must be retraced.

B The benchmark for comparison

A 'controversial issue' on the present law is whether, when evaluating an agreed compensatory damages clause, a court must compare the relevant amount with A's anticipated *actual* loss, or

---

161 See similarly Edwin Peel, 'Unjustified Penalties or an Unjustified Rule Against Penalties?' (2014) 130 LQR 365, 367, 370.

162 *Euro London Appointments Ltd v Claessens International Ltd* [2006] EWCA Civ 385, [2006] 2 Lloyd's Rep 436 [19]-[23].

163 *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [121].

164 *Lancore Services Ltd v Barclays Bank Plc* [2008] EWHC 1264 (Ch), [2008] 1 CLC 1039 [100], [102] affd [2009] EWCA Civ 752, [2010] 1 All ER 763.

her anticipated *legally-recoverable* loss.[165] Put differently, does the benchmark for comparison include loss that is legally irrecoverable? This section answers this question in three parts. The first part will set out the two views on this issue. The second part will examine the debate in the context of the remoteness doctrine. The final part will analyse the issue in the context of the mitigation doctrine.

I Two views

The literature contains two views on the issue of the proper comparator for any test of validity to adopt. They may be labelled the 'disassociated' and 'associated' view.

The disassociated view sees a distinction between the general damages and agreed damages regimes. It is said that the claimant may recover as general compensatory damages only her *legally-recoverable* loss. Conversely, the claimant may recover as liquidated compensatory damages her anticipated *actual* loss. Accordingly, when assessing whether a clause is valid, a court compares the agreed compensatory damages amount with A's prospective actual loss: that is, the prima facie factual loss that A might suffer, ignoring the effect of any general law damages rules. It follows that courts may include in the total benchmark sum a component for A's losses which would be legally-irrecoverable if A claimed general compensatory damages.

The crux of this view is that the test of validity ignores the remoteness and mitigation doctrines. An example will help. If A sued B at general law, she would be unable to recover for losses that were too remote, as the remoteness doctrine would exclude them, and for losses arising from her unreasonable conduct, as that would contravene the mitigation doctrine. Conversely, when applying the test of validity, the benchmark for assessment legitimately may include an amount representing either type of unrecoverable loss.

Many judges and jurists apparently take the disassociated view. For example, in *AMEV*, Mason and Wilson JJ held that agreed compensatory damages clauses enable A to provide for compensation in situations where the loss 'may not be recoverable at common law'.[166] Similarly, Rowan contends that agreed remedies properly seek 'to achieve a solution

---

165 Andrew Burrows (ed), *English Private Law* (3rd edn, OUP 2013) [21.180].

166 *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 193.

that is unavailable at law', such that the benchmark for comparison rightly includes 'loss which would not be compensable in the absence of the [agreed] damages clause'.[167]

Compare the associated view. This sees the general damages and agreed damages regimes as interconnected. Under the general law, a claimant may only recover for legally-recoverable loss. On this view, any agreed damages regime is merely an analogue of a general damages regime. Accordingly, under any agreed damages clause, the claimant may only recover for legally recognisable losses.

There exists strong support for the associated view. For instance, in *Ringrow*, the High Court held that a judge must compare 'what would be recoverable as [general] damages with the sum of money stipulated as payable on breach'.[168] Equally, in *Lansat*, Blair J held that the proper yardstick for comparison was anticipated legally-recoverable, not actual, loss.[169]

On the disassociated view, the comparative basis under any test of validity is anticipated actual loss. It follows that, in the agreed damages context, the remoteness and mitigation doctrines are simply ignored. In other words, the remoteness and mitigation questions are *not even asked*. However, the following will illustrate that the associated view is in fact correct. Correctly understood, the remoteness and mitigation doctrines *in principle* apply in the agreed damages context. However, the two rules *in fact* apply in such a way that, in most cases, they will not 'cut back' on the prospective loss that A may suffer at B's hands.[170] Put differently, the remoteness and mitigation questions are always asked, but in most cases they are *answered in the same way*. The answer is that no anticipated loss is too remote; and that no anticipated loss will result from A's unreasonable action. Consequently, the benchmark amount is unaffected. In most cases, then, a perceived distinction between a standard of prospective actual loss and of prospective legally-recoverable loss is illusory, as the former collapses into the latter.

II Remoteness

[167] Solene Rowan, 'For the Recognition of Remedial Terms Agreed Inter Partes' (2010) 126 LQR 448, 456, 469. See also Janet O'Sullivan, 'Lost On Penalties' (2014) 73 CLJ 480, 483.

[168] *Ringrow Pty Ltd v BP Australia Pty Ltd* [2005] HCA 71, (2005) 224 CLR 656 [21].

[169] *Lansat Shipping Co Ltd v Glencore Grain BV (The Paragon)* [2009] EWHC 551 (Com Ct), [2009] 2 All ER (Comm) 12 [22], affd [2009] EWCA Civ 855, [2010] 1 All ER (Comm) 459. See also *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [83].

[170] For a similar occurrence as regards the remoteness doctrine, see Edwin Peel, *The Law of Contract* (13th edn, Sweet and Maxwell 2011) [20-005] fn 25.

Generally speaking, the remoteness doctrine establishes the limits beyond which a defendant's responsibility for loss will not extend. As traditionally understood, the doctrine contains a 'reasonable foreseeability' test.[171] It comprises two overlapping rules. Under the 'first rule', a claimant may recover for a loss if it was the natural consequence of the breach. Under the 'second rule', the claimant may recover for a loss if, at the contract date, it was within the parties' reasonable contemplation that she would likely sustain that type of loss as a result of the defendant's breach. To recover for an unusual loss under the second rule, the defendant must be aware of the special circumstances exposing the claimant to the particular loss at or before the contract date.[172] Otherwise, the loss is legally irrecoverable.

On the issue of whether the remoteness doctrine is relevant to the test of validity, most leading scholars take the disassociated view. It is said that the doctrine is ignored.[173] Every writer cites *Robophone* as authority for the proposition. The case is conventionally analysed as follows. In the general damages context, to recover for any unusual loss under the second rule, A would be required to show that B knew of her special circumstances. Usually, A would prove that she communicated the relevant facts to B. However, in *Robophone*, Diplock LJ held that, in the agreed damages context, B need not know of the special circumstances likely to cause A the increased loss.[174] Accordingly, his Lordship considered that the yardstick for assessment properly could include loss that was beyond B's reasonable contemplation, and would therefore be too remote under the reasonable foreseeability test. It is said to follow that the remoteness rule is irrelevant to the test of validity; it simply 'does not need to be considered'.[175]

The traditional analysis, based on the dissociated view, is problematic. The reason is this: in *Robophone*, Diplock LJ actually supported a standard of anticipated *legally-recoverable* loss. His Lordship held that the agreed damages clause was to be tested against the damages that the plaintiffs would have recovered 'at common law, apart from that condition'.[176] Further, in *Makdessi*, the Court of Appeal interpreted *Robophone* as holding that 'a clause might be enforceable if the sum for which it provided was *justifiable* in the light of damage that would be

[171] *Hadley v Baxendale* (1854) 9 Exch 341, 354; 156 ER 145, 151.

[172] *Hadley v Baxendale* (1854) 9 Exch 341, 354-55; 156 ER 145, 151.

[173] eg Ewan McKendrick, *Contract Law* (10th edn, Palgrave Macmillan 2013) [21.5].

[174] *Robophone Facilities Ltd v Blank* [1966] 1 WLR 1428 (CA) 1448.

[175] JW Carter and Elisabeth Peden, 'A Good Faith Perspective on Liquidated Damages' (2007) 23 JCL 157, 171.

[176] *Robophone Facilities Ltd v Blank* [1966] 1 WLR 1428 (CA) 1442. See also 1446-47.

recoverable under the [remoteness doctrine]'.[177] In short, Diplock LJ viewed himself as applying the remoteness doctrine; and the Court of Appeal considered that one could rationalise how the test of validity operates in terms of the remoteness doctrine.

This confusion stems from a misreading of Diplock LJ's judgment. Once the case is properly read, the apparent anomalies disappear. His Lordship's reasoning in fact proceeded in three steps.[178] First, correctly understood, the second rule is based on the concept of the defendant assuming responsibility for the claimant's loss, not the parties' reasonable contemplation. Second, in the general damages context, a defendant *impliedly* undertakes to be responsible for the claimant's losses of a particular type arising from the breach. The defendant's knowledge of the special circumstances is merely a fact from which the undertaking can be inferred. Finally, in the agreed damages context, by agreeing to the clause B *expressly* undertakes to be responsible for all of A's losses of a particular type which the breach causes and which fall within the provision's scope, up to and including the agreed damages amount. Accordingly, notice of special circumstances is not required simply because B's assumption of responsibility is proved directly, rather than indirectly.

This analysis illustrates two points. For one, in principle, the remoteness doctrine applies in the agreed damages context. For another, *on the facts of every case*, A's anticipated loss will not be too remote: by the express clause, B will have voluntarily accepted responsibility for A's prospective loss. Consequently, the baseline amount will not be affected. It follows, then, that agreed compensatory damages are not an exception to the remoteness doctrine. Rather, they reveal a principled application of it.

Later cases in Australia and England have supported Diplock LJ's analysis in *Robophone*. As to the former, in *Stuart Pty Ltd v Condor Commercial Insulation Pty Ltd*, the New South Wales Court of Appeal followed the first two steps of Diplock LJ's reasoning in *Robophone*.[179] As to the latter, in *Transfield Shipping Inc v Mercator Shipping Inc (The Achilleas)*, Lord Hoffmann and Lord Hope reinterpreted the remoteness doctrine in English law.[180] Their Lordships held that, properly understood, the doctrine applies an 'assumption of responsibility' test. Under it, one must ask whether or not the defendant has objectively accepted responsibility for the kind, or

---

177 *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [79] (emphasis added).

178 *Robophone Facilities Ltd v Blank* [1966] 1 WLR 1428 (CA) 1448.

179 [2006] NSWCA 334, (2006) Aust Contract R 90-245 [53]-[54], [117].

180 [2008] UKHL 48, [2009] 1 AC 61.

type, of loss in question.[181] Peel contends that Lord Hoffmann and Lord Hope's judgments 'reinforce'[182] Diplock LJ's reasoning in *Robophone*. The judgments do so in two ways.

On the one hand, Lord Hoffmann and Lord Hope held that one factor indicating that a defendant has *not* assumed responsibility for a loss is where, as at the contract date, the loss is out of the defendant's control and is unquantifiable.[183] Conversely, in the case of agreed compensatory damages clauses, the amount of A's prospective loss necessarily will be quantified, as quantifiability is a central feature of this type of clause.[184] This suggests Diplock LJ was correct, and that an effective acceptance of responsibility will exist in every case.

On the other hand, in *Robophone*, Diplock LJ adopted assumption of responsibility reasoning in seeking to rationalise the second rule. However, the facts of *The Achilleas* concerned the first rule.[185] Accordingly, Lord Hoffmann and Lord Hope's judgments attempted to explain the *entire* doctrine using the concept that Diplock LJ had thought was only relevant to the second rule. Hence, it represented a *reinterpretation* of the remoteness doctrine though a natural extension.

Since the House of Lords decided *The Achilleas*, the English Court of Appeal has apparently accepted that the assumption of responsibility test represents English law.[186] Conversely, Australian courts are yet to consider the assumption of responsibility test. Hence, it is presently unclear whether Australian courts will accept it. That said, leading commentators suggest that the assumption of responsibility approach to the remoteness inquiry 'might portend a re-examination of extant contract law doctrine in Australia'.[187] Further, support for the assumption of responsibility test exists, namely the *Stuart* decision.

There is one final point. The assumption of responsibility approach to remoteness locates the intellectual underpinning of the doctrine in the parties' consensual relationship. The focus is entirely on that relationship; all considerations external to it are irrelevant.

---

[181] *The Achilleas* [2008] UKHL 48, [2009] 1 AC 61 [9], [12], [15], [21], [23], [30]-[32].

[182] Edwin Peel, *The Law of Contract* (13th edn, Sweet and Maxwell 2011) [20-136] fn 841.

[183] *The Achilleas* [2008] UKHL 48, [2009] 1 AC 61 [23], [34], [36]. See further Leonard Hoffmann, '*The Achilleas*: Custom and Practice or Foreseeability?' (2010) 14 Edin LR 47, 59.

[184] Text to n 13 in ch 1.

[185] *The Achilleas* [2008] UKHL 48, [2009] 1 AC 61 [6], [93].

[186] *Supershield Ltd v Siemens Building Technologies FE Ltd* [2010] EWCA Civ 7, [2010] 2 All ER (Comm) 1185 [43]; *John Grimes Partnership Ltd v Gubbins* [2013] EWCA Civ 37, [2013] BLR 126 [24].

[187] Nicholas Seddon, Rick Bigwood and Manfred Ellinghaus, *Cheshire and Fifoot Law of Contract* (10th Aus edn, LexisNexis Butterworths 2012) xii.

Consequently, the assumption of responsibility approach is faithful to corrective justice norms. Chapter 8 explains corrective justice, and why it justifies the Penalty Rule.

III Mitigation

Broadly speaking, the mitigation doctrine concerns the steps that a claimant has, or should have, taken to reduce or avoid her loss. The doctrine comprises various rules.[188] For present purposes, only one rule is relevant. A claimant must take all reasonable steps to minimise or avoid the loss that a defendant's breach causes. This is often called the 'duty to mitigate'. The 'avoidable loss' rule holds that the claimant cannot recover general damages for any consequential loss that she may have reduced or avoided by taking reasonable steps; that is, where she has failed to comply with her 'duty to mitigate'. For simplicity, in this thesis, the 'mitigation doctrine' solely refers to the avoidable loss rule, unless the context otherwise requires.

Three introductory points arise. To start, the 'duty to mitigate' label is misleading. Once a claimant has been wronged, she is under no *duty* to minimise her losses, for the defendant has no corresponding right that the claimant takes action. Next, the theoretical basis of the 'duty' to mitigate is difficult. As orthodoxly—and best—understood, it rests on causation.[189] That is, the claimant's failing reasonably to take steps to minimise or avoid the relevant loss constitutes intervening conduct that breaks the causal chain, such that the claimant's acts have caused her own loss.[190] Lastly, the traditional, causation-based theory of the mitigation doctrine views it as concerned with determining whom, as between the wrongdoer and victim, is properly responsible for loss on its face sustained as a result of the defendant's wrong. The relevant inquiry, then, exclusively focuses upon the parties' relationship; and seeks to do justice as between the parties. Accordingly, this explanation of the doctrine is consistent with corrective justice theory. Chapter 8 explains what this theory entails.

On the issue of whether the mitigation doctrine is relevant to the test of validity, some academics support the associated view. For example, Beale states that the baseline for comparison includes A's 'net loss' after taking account of A's expected ability to mitigate her

[188] Harvey McGregor, *McGregor on Damages* (19th edn, Sweet and Maxwell 2014) [9-003]-[9-006]; *Cardno BSD Pty Ltd v Water Corporation (No 2)* [2011] WASCA 161 [145].

[189] *Standard Chartered Bank v Pakistan National Shipping Corp* [2001] EWCA Civ 55, [2001] 1 All ER (Comm) 822 [21], [41]; *Castle Constructions Pty Ltd v Fekala Pty Ltd* [2006] NSWCA 133, (2006) 65 NSWLR 648 [19]-[21].

[190] *British Westinghouse Electric and Manufacturing Co Ltd v Underground Electric Railways Co of London Ltd* [1912] AC 673 (HL) 689.

loss.[191] However, most prominent scholars take the disassociated view. They contend that the doctrine is ignored.[192] These academics usually cite two English Court of Appeal cases in support: *Abrahams v Performing Right Society Ltd*[193] and *Murray*.

As to the former, in *Abrahams*, Hutchison LJ seemingly held that arguments regarding the 'duty' to mitigate were irrelevant where A claims against B under an agreed damages clause. In his Lordship's words, 'the concept of a duty to mitigate is entirely foreign to a liquidated damage[s] claim'.[194]

As to the latter, in *Murray*, at first instance Burnton J had held the agreed damages clause was penal because it took no account of Murray's 'duty' to mitigate.[195] Conversely, the English Court of Appeal held the clause was valid. Courts and commentators have interpreted the Court of Appeal as upholding the clause in spite of the fact that it took no account of the 'duty' to mitigate.[196] Undeniably, some support for this view exists. In *Murray,* Arden LJ held that the clause did not make any allowance for Murray's 'duty' to mitigate, as it did not include a credit for remuneration he could reasonably have earned from other sources in the relevant period.[197] As a result, it is said that the test of validity disregards the mitigation doctrine.

Thus, the leading authorities seemingly support the view that the benchmark for assessment ignores the mitigation doctrine. Not so. The law is best understood otherwise. There are five points.

First, in principle, the mitigation doctrine applies in the agreed damages context. That is, in every case, the parties are required to ask the following question: given A's prospective loss, what part of that loss could A reasonably be expected to minimise or avoid? The decisions in *Abrahams* and *Murray* seemingly stand in the way of this analysis. Properly interpreted, they do not.

As to *Abrahams*, Hutchison LJ is commonly thought to have held that the 'duty' to mitigate is irrelevant to the benchmark for assessment. However, his Lordship's statements were in fact directed to whether the mitigation doctrine applies in the context of a *liquidated*

---

[191] Hugh Beale (gen ed), *Chitty on Contracts* (2nd supp, 31st edn, Sweet and Maxwell 2013) [26-175].

[192] eg Harvey McGregor, *McGregor on Damages* (19th edn, Sweet and Maxwell 2014) [15-022].

[193] [1995] ICR 1028 (CA).

[194] *Abrahams v Performing Right Society Ltd* [1995] ICR 1028 (CA) 1040.

[195] *Murray v Leisureplay plc* [2004] EWHC 1927 (QB) [102]-[104].

[196] eg *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [92], [95]; John McGhee (gen ed), *Snell's Equity* (33rd edn, Sweet and Maxwell 2014) [13-018].

[197] *Murray v Leisureplay plc* [2005] EWCA Civ 963, [2005] IRLR 946 [65], [69], [73].

compensatory damages clause (that is, an agreed damages clause that a court has upheld). His Lordship's view that the mitigation doctrine is irrelevant in that context accords with the argument, put forward in Chapter 6, that monetary liquidated damages are a debt. However, the decision cannot be taken as suggesting that the 'duty' to mitigate is irrelevant at the anterior stage of assessing whether an agreed compensatory damages clause is valid, or not. Two examples will illustrate the point.

For one, the reasoning behind Hutchison LJ's comment that the 'duty' to mitigate was 'entirely foreign' was this: 'How could it be right to hold [A], who can show that [her incurred loss] is *greater*, to the stipulated sum, but permit [B] who can show that it is less to take advantage of that fact? Why should such an … inconsistent approach be approved …'.[198] As Chapter 6 demonstrates, generally where A's incurred loss is more than the *liquidated* compensatory damages amount, A can nevertheless only recover that amount. Accordingly, the context in which the judge was comparing the two approaches was where the relevant clause is valid.

For another, his Lordship held that 'where there is a liquidated damage clause which is valid—[ie] cannot be impugned as a penalty—there is no room for arguments on mitigation of damages …'.[199] Put differently, if A sues B to recover a sum under a liquidated compensatory damages clause, the mitigation doctrine does not apply. Whether she *in fact* failed to minimise or avoid any losses arising after the breach date is irrelevant.

As to *Murray*, subsequent courts have interpreted the Court of Appeal as holding that the 'duty' to mitigate is ignored when calculating the baseline amount.[200] However, that is not quite right. In fact, Buxton LJ held that, in principle, the parties must 'take account' of the 'duty' to mitigate when calculating any agreed compensatory damages amount—by at least asking the mitigation question—although doing so need not necessarily affect the agreed damages sum.[201] We will return to this point in a moment. Notably, Clarke LJ agreed with Buxton LJ's approach.

Conversely, Arden LJ seemingly held that the agreed damages clause did not include a proper discount for Murray's anticipated mitigation opportunities. However, as her Ladyship acknowledged, *on the facts* no significant discount applied: the restrictive covenants applying to

---

198 *Abrahams v Performing Right Society Ltd* [1995] ICR 1028 (CA) 1040 (original emphasis).

199 *Abrahams v Performing Right Society Ltd* [1995] ICR 1028 (CA) 1041.

200 *Tullett Prebon (Australia) Pty Ltd v Purcell* [2009] NSWSC 1079 [126], affd [2010] NSWCA 150.

201 *Murray v Leisureplay plc* [2005] EWCA Civ 963, [2005] IRLR 946 [115].

Murray following termination effectively meant that any prospective mitigation opportunities were of minimal value. In Arden LJ's words, the contract contained 'some fierce restrictions on competition', which covenants 'would undermine any point that [could] be made about [Murray] being … well able to obtain employment … if he was summarily dismissed by MFC'.[202] Accordingly, her Ladyship's judgment is best interpreted as follows. In principle, the 'duty' to mitigate applied. However, on the facts, the restrictions on competition following termination would prevent Murray from effectively mitigating, undermining any substantial future mitigation opportunities. Consequently, the benchmark for comparison was not significantly affected.

This analysis of *Murray* explains why judges and jurists have mistakenly thought that the Court of Appeal ignored the mitigation doctrine in applying the test of validity.

Second, an important distinction exists between assessing A's prospective *loss* and assessing her anticipated *mitigation opportunities*. For three reasons, pre-assessment is more difficult in the latter case. To start, the reasonable steps required of A to minimise her loss will depend entirely on the precise nature and extent of her prospective loss. Accordingly, A's anticipated mitigation opportunities are one step removed from her prospective loss. Next, in calculating A's future mitigation opportunities, the parties will need to consider additional variables, such the anticipated behaviour of third parties. Lastly, A will likely hold the relevant information grounding the assessment of her prospective loss. Conversely, A may be unaware of many critical facts relevant to determining her anticipated mitigation opportunities.

Third, agreed damages clauses typically deal with low-probability, temporally distant events. In most cases, it is exceedingly difficult to predict, at the contract date, how B's breach will affect A's factual position. Given the distinction previously discussed, in such cases it is 'unrealistic' to expect the parties prospectively to identify A's mitigation opportunities.[203] It follows that, in these circumstances, one simply cannot know what future conduct will be reasonable or unreasonable in response to the contemplated breach. To explain, what will constitute reasonable post-breach conduct will depend heavily on the particular circumstances in which A will be placed. If the parties cannot clearly specify these circumstances, a court simply cannot say that *any* prospective act of A in the wake of B's wrong will be unreasonable. It follows that no anticipated conduct of A can be classed as 'unreasonable', and no

---

[202] *Murray v Leisureplay plc* [2005] EWCA Civ 963, [2005] IRLR 946 [71].

[203] Andrew Burrows, *Remedies for Torts and Breach of Contract* (3rd ed, OUP 2004) 446.

prospective loss will result from A's intervening acts. The baseline amount will not therefore be affected.

Thus, in most cases, the mitigation doctrine in principle applies, but *on the facts* it will not affect the outcome of A's claim on an agreed compensatory damages clause. Put differently, the mitigation question is asked, but usually it is answered in such a way that it does not affect how the benchmark amount is calculated.

Fourth, Buxton LJ's reasoning in *Murray* supports this analysis. As best interpreted, his Lordship held as follows.[204] In principle, the parties must 'take account' of the 'duty' to mitigate when calculating any agreed compensatory damages sum; they must ask the mitigation question. However, in hard cases where the parties cannot sensibly cater for the 'duty' to mitigate, they may 'take account' of it simply by considering the issue; the agreed damages sum need not *necessarily* be affected. In other words, where the parties cannot intelligibly pre-determine A's prospective mitigation opportunities, correctly applied the mitigation doctrine will not impact on the benchmark for assessment.

Finally, in most cases, assessed as at the contract date, it will be exceedingly difficult to predict the effects of B's contemplated breach, and A's mitigation opportunities will be totally obscure. However, in *some* cases, the parties will be able more precisely to divine the consequences of B's wrong, and A's opportunities will be sufficiently clear. If so, in principle the mitigation doctrine applies *and* on the facts it will affect the baseline for assessment. One example[205] will have to suffice.[206]

Cine Bes Filmcilik ve Yapimcilik AS v United International Pictures involved a licence to exhibit films on a pay movie channel.[207] The agreement stipulated that if the licensee failed to pay its license fees, the licensor obtained the right to terminate the contract; the licensee was obligated to pay various amounts; and the licensee's rights to any currently licensed films were to 'terminate and revert' to the licensor. The clause did not require the licensor to give any credit to the licensee for the value of the rights to use the films with current licences that it 'recovered' (that is, the value of the ability to exploit those rights for the balance of the licence periods). The licensee defaulted; the licensor terminated. The issue before the English Court

---

[204] See similarly Edwin Peel, *The Law of Contract* (13th edn, Sweet and Maxwell 2011) [20-136].

[205] See also *Cedar Meats Pty Ltd v Five Star Lamb Pty Ltd* [2013] VSC 164 [108], revd [2014] VSCA 32.

[206] One may classify this issue as relating to 'compensating advantages' or the mitigation doctrine. For example, see *Cine Bes Filmcilik ve Yapimcilik AS v United International Pictures* [2003] EWCA Civ 1669, [2004] 1 CLC 401 [29] (issue of 'countervailing benefits'), [53] (mitigation issue). Nothing substantive turns on this. In particular, both doctrines rest on causation.

[207] [2003] EWCA Civ 1669, [2004] 1 CLC 401.

of Appeal was whether the licensee's argument that the agreed damages clause was penal had a real prospect of success.

The Court held that a triable issue existed. The agreed damages clause was arguably penal because it did not provide the licensee with any credit for the benefit of the film rights 'recovered' on termination, being the value of the film rights for the balance of the licence periods. Thomas LJ held that the benefit to the licensor arising from the termination was 'the opportunity to exploit the balance of the licence period of films for which [the licensee] had paid' by licensing the films to others.[208] This benefit was one 'arising in the ordinary course of business as a consequence of the termination'; a benefit that would 'at first sight have been obvious to the parties at the time the contract was made'.[209] Accordingly, the parties were required to bring into account that benefit when determining the agreed damages sum, and they did not. In Mance LJ's words, the clause failed 'to address the countervailing benefits which would, on their face, appear likely … as flowing from any termination …'.[210]

In summary, many judges and jurists contend that the remoteness and mitigation doctrines do not apply at all in the agreed damages context. It is said that courts simply ignore these doctrines. In truth, the rules apply, but in such a way that they do not affect the benchmark amount. The courts' treatment of the basis for assessment merely reflects a proper understanding of how the remoteness and mitigation doctrines operate in the agreed damages context.

C The assessment date

It is trite law that the tests of validity are applied as at the contract date, not a later date, such as the date of the breach, or the time of judgment.[211] In other words, the validity of an agreed damages clause is judged *ex ante*, not *ex post*. Appellate courts in England and Australia have continually confirmed this.[212]

In principle, a contract date assessment date provides the maximum possible respect for the parties' promises short of abolishing the Penalty Rule completely. This is because if a

[208] *Cine Bes Filmcilik ve Yapimcilik AS v United International Pictures* [2003] EWCA Civ 1669, [2004] 1 CLC 401 [51].

[209] *Cine Bes Filmcilik ve Yapimcilik AS v United International Pictures* [2003] EWCA Civ 1669, [2004] 1 CLC 401 [47], [50].

[210] *Cine Bes Filmcilik ve Yapimcilik AS v United International Pictures* [2003] EWCA Civ 1669, [2004] 1 CLC 401 [29].

[211] eg *Comr of Public Works v Hills* [1906] AC 368 (PC) 376; *Hamilton v Lethbridge* (1912) 14 CLR 236 (HCA) 264. cf *Unaoil Ltd v Leighton Offshore Pte Ltd* [2014] EWHC 2965 (Com Ct) [71] (date of a significant amendment).

[212] eg *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [42]; *Spiers Earthworks Pty Ltd v Landtec Projects Corporation Pty Ltd (No 2)* [2012] WASCA 53, (2012) 287 ALR 360 [21].

court was permitted to take into account *ex post* events, this might provide a wider basis upon which to impugn an agreed damages clause, and hence a greater potential to undermine the parties' agreement.

The principle is clear enough. However, on occasion, it has been misapplied. Consider *Ringrow*. Ringrow and BP entered into three agreements. First, pursuant to a sale contract, Ringrow agreed to buy from BP the proprietary right to a service station. The purchase price *included* goodwill. Second, under a franchise agreement entitled the BP Branded Privately Owned Sites Agreement—the 'POSA'—Ringrow agreed exclusively to purchase fuel supplies from BP for five years. If Ringrow breached any term of the POSA, BP was entitled to terminate the agreement. Finally, pursuant to an option deed, if BP exercised its right to terminate the POSA, it obtained an option to re-acquire the proprietary right to the service station at its market value, but the figure was to *exclude* goodwill. Ringrow breached the POSA, purchasing fuel from another supplier. BP terminated the POSA. BP then sought to exercise its option to re-purchase. Ringrow contended that the option clause was penal. One of its arguments rested on the fact that the resale price excluded goodwill. However, the only expert evidence adduced regarding goodwill proved that no valuable goodwill existed as at the re-transfer date.

The Australian High Court unanimously held that the clause was valid. In reaching its decision, the High Court held that the *Dunlop* judgment represented the law in Australia.[213] It follows that, in principle, the Court should have compared BP's anticipated loss flowing from Ringrow's breach, and the anticipated agreed damages value that Ringrow would be obliged to transfer. So, on the facts, the re-purchase price constituted the relevant 'loss'; and the value of the proprietary right to the land and improvements constituted the agreed damages value. Accordingly, the High Court should have compared the anticipated *future* re-purchase price and the anticipated *future* value of the proprietary right to the site, assessed as *at the contract date*. In particular, to assess correctly the anticipated value of the proprietary right, the Court should have inquired into, and taken account of, the estimated goodwill value as at the anticipated future re-transfer date, calculated as at the contract date.

Conversely, the High Court compared the re-purchase price *in fact paid* with the value of the proprietary right *in fact transferred*, assessed as *at the re-transfer date*.[214] In particular, as a result of its finding that no valuable goodwill existed as at the re-transfer date, the Court

[213] *Ringrow Pty Ltd v BP Australia Pty Ltd* [2005] HCA 71, (2005) 224 CLR 656 [12].

[214] *Ringrow Pty Ltd v BP Australia Pty Ltd* [2005] HCA 71, (2005) 224 CLR 656 [21].

decided that in assessing the value of the proprietary right, it would allocate no amount to goodwill. In short, the High Court applied the test of validity *ex post*, and not *ex ante*, as required.

In *Zachariadis*, the Victorian Court of Appeal purported to rationalise the anomaly on two bases.[215]

On the one hand, the Court suggested that the difference arose because the obligation was to transfer a proprietary right rather one to pay money. This contention is specious. In this context, the private law cannot rationally distinguish between monetary and non-monetary entitlements. The question is whether a court can *value* the subject matter of the particular obligation that the agreed damages clause triggers. Usually, the obligation is one to pay money, such that the value is self-evident. However, where the obligation is to transfer a proprietary right, a court is able to value that right, if necessary by expert evidence. Accordingly, no proper basis for different treatment exists.

On the other hand, the Court of Appeal suggested that the High Court might have regarded the goodwill valuation as casting light on the circumstances existing at the time of the contract. Some authorities support the view that events occurring after formation provide circumstantial evidence as to what the parties, as at the contract date, reasonably could have expected to occur.[216]

However, at best, any evidence of *ex post* events has little probative value, as it does not reliably indicate what the parties reasonably expected as at the contract date. Whether a particular loss in fact materialised or not reveals nothing, by itself, about whether the parties reasonably anticipated it at the contract date. In *Zachariadis*, the Court of Appeal acknowledged this.[217] Ironically, therefore, the Court's own judgment cuts against its purported explanation of the High Court's judgment in *Ringrow*.

Further, at worst, taking into account evidence of *ex post* events is dangerous. If a court admits the evidence, it might influence the judge to persuade herself that the parties contemplated the eventuality merely because it occurred, however remote the possibility was at the contract date. The law can minimise this danger by excluding evidence of post-formation

---

215 [2009] VSCA 258, (2009) 26 VR 47 [154].

216 eg *Philips Hong Kong Ltd v A-G of Hong Kong* (1993) 61 BLR 41 (PC) 59.

217 *Zachariadis v Allforks Australia Pty Ltd* [2009] VSCA 258, (2009) 26 VR 47 [154].

events for the purpose of proving that the clause is valid. Accordingly, evidence of events occurring after the contract date is properly 'irrelevant and inadmissible'.[218]

In sum, the assessment date for the test of validity seems like a minor issue. In fact, it is critical. A failure to understand its importance is what has led many to misunderstand the Penalty Rule, and its true relationship with the general damages regime. Chapter 7 considers the issue.

3 Conclusion

This chapter analysed two key questions, and three pressing problems, for the Penalty Rule tests of validity.

The first question was whether each test of validity still works. The chapter illustrated that, in principle, the Estimation Test is problematic for at least three reasons: its true interpretation is uncertain; it leads to incoherence; and it lacks fit. Further, the chapter showed that the Greater Sum Test no longer works; and that the Indiscriminate Cause Test remains at a crossroads. As for the Impossibility Test, a loss-based analysis suggests that it should be jettisoned on the basis that it is incoherent. But the test can be saved if the law embraces a rights-based approach. Lastly, the chapter demonstrated that the Extravagance Test works well, and exhibits great explanatory value.

The second question was put thus: what is the primary, or sole, test of validity? The chapter illustrated that the conventional view in England and Australia is that the Estimation Test is primary. However, it also showed that both jurisdictions have moved from this stance. As a matter of precedent, the Extravagance Test is now the sole test applied. Anglo-Australian law has worked itself pure.

This chapter then turned to consider three issues that have plagued the test of validity. It showed that the Commercial Justification Test exhibits a serious collapse in legal understanding. The chapter then revealed that a proposed distinction between a standard of anticipated actual loss and anticipated legally-recoverable loss is illusory. In closing, it outlined how a court is properly to apply a test of validity that operates prospectively, and assesses matters as at the contract date.

[218] *Clydebank Engineering and Shipbuilding Co Ltd v Don Jose Ramos Yzquierdo y Castaneda* [1905] AC 6 (HL) 17.

6 CONSEQUENCES

There are many cases regarding the consequences of applying the test of validity. The issues they raise are intricate. Several cases appear contradictory. This chapter attempts to bring order to the chaos.

This chapter examines consequential issues. It is divided into two sections. The chapter begins by examining the consequences of a court upholding an agreed damages clause as a liquidated damages clause. It focuses specifically on the nature of liquidated damages; on the situation where A's incurred loss is *less* than the liquidated compensatory damages amount; and the scenario where A's incurred loss *exceeds* that amount. The chapter then considers three issues concerning invalid penalty clauses. First, it analyses the true nature of a penalty, including whether a penalty clause is properly unenforceable or void. Second, the chapter investigates the effect of such a clause when A's loss is *less* than the penalty amount. Finally, it evaluates how a penalty clause operates when A's loss *exceeds* that amount.

The chapter reveals much about each type of clause. On the one hand, properly interpreted, liquidated damages clauses are relatively straightforward. A rights-based approach solves the most pressing issues. On the other hand, penalty clauses raise several puzzles that have remained unsolved for decades. However, recent case law has rediscovered the answers. The resulting picture of the Penalty Rule is one of consistency, symmetry and logicality: in other words, a very attractive one.

1 Liquidated damages clauses

A Nature

In the paradigm scenario,[1] an agreed damages clause imposes upon B an obligation to pay money. If the clause is valid, it is a liquidated damages clause. If so, A obtains a right to enforce the provision against B according to its terms. This section addresses one question: properly understood, in the 'core case' are liquidated damages in the nature of general damages, or in the nature of a debt? The question is simple. The answer is not.

One must begin with the concept of a debt. According to Beale, a debt is 'a definite sum of money fixed by the agreement' that one party must pay in return for the other party

[1] Text following n 20 in ch 1.

performing an obligation, or upon some specified event occurring.[2] The relevant amount must be certain, in the sense of being ascertainable when payment is due. Relevantly, a debt displays five features. First, a claimant seeking payment of a debt need only prove that she performed the obligation, or that the relevant event occurred; she need not prove that she incurred any loss in order to recover.[3] Second, the remoteness doctrine is 'irrelevant'.[4] Third, the mitigation doctrine 'does not generally apply'.[5] Fourth, the claimant may seek summary judgment. Finally, the claimant will become an unsecured creditor upon the defendant's insolvency.

The highest authority in England and Australia has held that liquidated damages are not in the nature of a debt. In *President of India v Lips Maritime Corp (The Lips)*, Lord Brandon held that a form of liquidated compensatory damages known as 'demurrage' is not 'a liability sounding in debt', but rather:

> It is a liability in damages to which a charterer becomes subject because, by detaining the chartered ship beyond the stipulated lay days, he is in breach of his contract … The effect of [a demurrage] clause is to liquidate the damages payable: it does not alter the nature of the charterer's liability, which is and remains a liability for damages, albeit liquidated damages.[6]

The Australian High Court has endorsed this reasoning.[7] However, for two reasons, his Lordship's statement is challengeable.

On the one hand, Lord Brandon argues that demurrage is in the nature of damages, and not debt, because the liability arises upon a breach of contract. In rights language, a liquidated damages clause is in nature of damages because it involves a *secondary* obligation to pay, not a *primary* obligation to pay. It is true that a debt concerns a primary obligation. However, properly understood, so do liquidated damages. As Chapter 8 illustrates, correctly comprehended liquidated damages clauses impose a primary duty on B to pay the stipulated sum, conditional on B being in default. In short, B's obligation arises *upon* his default, but not *because of* his default.

---

[2] Hugh Beale (gen ed), *Chitty on Contracts* (2nd supp, 31st edn, Sweet and Maxwell 2013) [26-008].

[3] *Jervis v Harris* [1996] Ch 195 (CA) 202.

[4] *Jervis v Harris* [1996] Ch 195 (CA) 202.

[5] Hugh Beale (gen ed), *Chitty on Contracts* (2nd supp, 31st edn, Sweet and Maxwell 2013) [26-008] (footnote omitted).

[6] [1988] AC 395 (HL) 422. See also 424.

[7] *Hungerfords v Walker* (1989) 171 CLR 125 (HCA) 139, 162.

On the other hand, Lord Brandon held that the nature of B's liability under a liquidated damages clause is that of a general damages obligation. However, a claimant cannot enforce a right to general compensatory damages without litigation, as the quantum of the damages must be judicially assessed. Conversely, if B does not challenge the clause's validity, A may recover the liquidated compensatory damages sum without restoring to court proceedings.

Further, the proposition that liquidated damages are not in the nature of debts is unhistorical and unprincipled.

As to history, Chapter 2 noted that courts first started distinguishing between liquidated damages clauses and penalties clauses in the 18th century. At that time, the plaintiff recovered liquidated damages at law 'as a debt'.[8]

As to principle, conceptually liquidated damages are in the nature of a debt. Liquidated damages display each of the indicia of a debt. First, to recover liquidated compensatory damages, A need not only prove that B defaulted; A need not prove her incurred loss. Second, where A claims under a liquidated compensatory damages clause, 'no question of mitigation arises'.[9] Third, in principle, remoteness issues are irrelevant to such claims. Fourth, A may recover liquidated compensatory damages 'in much the same way as a debt', such that A may access summary disposition proceedings.[10] Finally, in principle, A becomes an unsecured creditor upon B's insolvency.

The conclusion is clear. A liquidated damages clause allows A to recover 'a liquidated sum due'.[11] That is a debt by another name. It follows that, properly viewed, monetary liquidated damages are in the nature of debts, not damages.

B Incurred loss *less* than liquidated compensatory damages

Assume that A's incurred loss is *less* than the liquidated compensatory damages amount. May she still enforce the liquidated compensatory damages clause?

The short answer is 'yes'. For instance, in *Ng v Ashley King (Developments) Ltd*, Lewison J held that A may recover from B 'the amount of the liquidated damages and only that amount.

---

8 Chris Rossiter, *Penalties and Forfeiture* (Lawbook Company 1992) 13.

9 *Abrahams v Performing Right Society Ltd* [1995] ICR 1028 (CA) 1041. See similarly *Cedar Meats Pty Ltd v Five Star Lamb Pty Ltd* [2013] VSC 164 [14], revd [2014] VSCA 32.

10 *Carter on Contract* [43-010] (service 37, April 2014).

11 *Metro-Goldwyn-Mayer Ltd v Greenham* [1966] 2 NSWR 717 (NSWCA) 722.

Whether [her incurred] loss is greater than or less than the contractual amount, [she] is entitled to the contractual amount.'[12]

Thus, incurred loss is irrelevant. It follows that A may recover the liquidated compensatory damages sum even if she sustained no loss at all, or actually benefitted from B's breach.

C Incurred loss *more* than liquidated compensatory damages

In some cases, A's incurred loss may be *more* than the liquidated compensatory damages sum. This situation will arise, for example, where the parties agree that B will be obligated to pay to A one sum upon a number of varying breaches occurring, and the most serious breach in fact occurs. Can A recover general compensatory damages in addition to, or instead of, liquidated compensatory damages? Put differently, does a liquidated compensatory damages clause provide an exhaustive remedy? The seemingly contradictory cases are reconcilable. Three steps are required. The first step is to outline how best to understand the present case law. The second step is to examine the apparently difficult authorities with that understanding in mind. The final step is to illustrate why the account of the law offered is the most principled.

I An account of the law

Properly understood, the best account of the law can be stated in four principles. The first is a general principle. The next three are subsidiary principles that follow from it.

First, liquidated compensatory damages clauses provide an exclusive remedy for compensatory damages for B's default. So, in *Dura (Australia) Constructions Pty Ltd v Hue Boutique Living Pty Ltd*, Dixon J held that 'The basic principle is that a valid liquidated damages clause will represent an exhaustive remedy for breach of the [relevant] obligation …'.[13] It is said that the authorities 'are at one':[14] a liquidated damages clause excludes general damages.

Second, A may not claim general compensatory damages *instead of* liquidated compensatory damages. Hence, in *Temloc Ltd v Errill Properties Ltd*, Croom-Johnson LJ held

---

12 [2010] EWHC 456 (Ch), [2011] Ch 115 [19]. See similarly *J-Corp Pty Ltd v Mladenis* [2009] WASCA 157, (2010) 26 BCL 106 [35].

13 [2012] VSC 99, (2013) 29 BCL 19 [644] (footnote omitted), affd [2013] VSCA 179. See similarly *Temloc Ltd v Errill Properties Ltd* (1987) 39 BLR 30 (CA) 39.

14 *Silent Vector Pty Ltd v Squarcini* [2008] WASC 246, (2009) 25 BCL 58 [62].

that where a liquidated compensatory damages clause exists, A cannot elect between claiming general and liquidated damages: 'there is no choice available'.[15]

Third, A cannot claim general compensatory damages *in addition to* liquidated compensatory damages. Thus, in *NLS Pty Ltd v Hughes*, Bawrick CJ held that a liquidated compensatory damages clause imposes 'a limitation upon the amount recoverable' as compensation for any relevant breach.[16] An English case often cited as illustrating this point is *Cellulose Acetate Silk Co Ltd v Widnes Foundry (1925) Ltd*.[17] A contractor agreed to build an acetone recovery plant for a principal. The relevant clause provided that, if the contractor delayed, it would pay to the principal a 'penalty' of £20 for each week of delay. The contractor completed 30 weeks late. The principal suffered losses of £5,850. The House of Lords held that the clause was valid, such that the principal was only entitled to £600. Accordingly, although the principal's incurred loss was much more than the liquidated compensatory damages sum, it was entitled to recover only the stipulated sum. It necessarily follows that the House of Lords viewed the liquidated damages clause as an exclusive remedy.[18]

Finally, any exclusivity only applies for the particular breach or breaches of obligation, and kinds of consequential loss, falling within the scope of the liquidated compensatory damages clause. It follows that, even if a liquidated damages clause exists, A may recover from B general compensatory damages for an independent kind of loss flowing from B's breach of a separate contractual obligation owed to A. However, for A to recover from B general compensatory damages in addition to liquidated compensatory damages, A must demonstrate *both* that B breached a separate contractual obligation owed to A; and that B's separate breach caused an independent kind of loss.[19] As to the former, A must establish that B breached a contractual obligation distinct from the obligations which the liquidated compensatory damages clause covers because that clause 'covers the field' regarding the compensatory response to a breach falling within its scope. As to the latter, A must prove that she incurred a kind of loss distinct from the kinds of losses which the liquidated compensatory damages clause covers because that clause is exhaustive concerning the kinds of consequential loss

---

[15] (1987) 39 BLR 30 (CA) 38. See similarly *J-Corp Pty Ltd v Mladenis* [2009] WASCA 157, (2010) 26 BCL 106 [35].

[16] (1966) 120 CLR 583 (HCA) 589.

[17] [1933] AC 20 (HL).

[18] See *Suisse Atlantique Société d'Armement Maritime SA v NV Rotterdamsche Kolen Centrale* [1967] 1 AC 361 (HL) 395-96.

[19] *Richco International Ltd v Alfred C Toepfer International GmbH (The Bonde)* [1991] 1 Lloyd's Rep 136 (Com Ct) 142. A breach of a separate duty is necessary, but not sufficient: *Inverkip Steamship Co Ltd v Bunge and Co* [1917] 2 KB 193 (CA) 199, 203.

falling within its scope. In short, A can recover general damages from B in addition to liquidated damages because the relevant obligation, and the kind of consequential loss, falls outside the provision's ambit, properly construed.

An example may help. Assume a building contract exists between A, the principal, and B, the contractor. It contains a liquidated compensatory damages clause stating that B is obliged to pay A £500 per week for the period of its delay. A cannot recover liquidated compensatory damages for consequential loss flowing from B's breach of its obligation to perform on time. However, B also undertakes a separate obligation as to quality. If B breaches that obligation, A may claim general compensatory damages for any consequential loss, for example the rectification costs.

II Examining the difficult case law

This account of the law presents a clear path. However, it might be suggested that it is misleading, in that two roadblocks exist. These are decisions seemingly involving *non-exhaustive* liquidated damages clauses. Properly understood, they do not.

The first apparent roadblock is *J-Corp Pty Ltd v Mladenis*.[20] Owners engaged a builder to construct a house on the owners' land. The builder prepared a standard form contract. Under clause 11.1, the builder was obliged to complete the works within 52 working weeks of commencement. Pursuant to clause 11.9, if the builder breached clause 11.1, it was liable to pay to the owners agreed damages 'at the rate of NIL DOLLARS ($00.00) per day for each day'[21] it delayed. The builder completed late. The owners argued that clause 11.9 merely ousted any claim for agreed compensatory damages, such that they were entitled to sue for general compensatory damages. Conversely, the builder argued that, given the clause, it was not liable to pay *any* damages for delay. The Western Australian Court of Appeal held that the owners were entitled to claim general compensatory damages for delay.

On one view, the decision suggests that liquidated compensatory damages clauses are not exclusive.[22] So, the agreed compensatory damages sum was 'nil', and the clause was valid, but nonetheless the owners were entitled to recover general compensatory damages.

---

[20] [2009] WASCA 157, (2010) 26 BCL 106.

[21] *J-Corp Pty Ltd v Mladenis* [2009] WASCA 157, (2010) 26 BCL 106 [14].

[22] Nicholas Seddon, Rick Bigwood and Manfred Ellinghaus, *Cheshire and Fifoot Law of Contract* (10th Aus edn, LexisNexis Butterworths 2012) [23.45] fn 367.

However, carefully read, this is not what the Court of Appeal said. In fact, the Court held that the contract did not contain an agreed damages clause *at all*. Buss JA held that, construed in the context of the whole contract, clause 11.9 displayed an intention that 'the standard form [agreed] damages clause should … be excluded'; for Newnes JA, the provision's wording made it clear that it was 'to have no effect and that the [owners were] to be left with the burden of proving such [loss] as they may be able to establish'.[23] In essence, no agreed damages clause existed. Accordingly, the owners had a subsisting right to claim general compensatory damages for any incurred loss.

The second roadblock is a particular line of charterparty cases. In these cases, it is said that the courts have permitted A to recover general compensatory damages in addition to liquidated compensatory damages. For example, according to Furst and Ramsey, these cases illustrate that a demurrage clause is 'not an exhaustive remedy for a breach giving rise to losses of more than one character'.[24]

Two definitional points must be addressed. The first is the meaning of a 'demurrage' clause. In a charterparty, a charterer undertakes a strict liability obligation to load or discharge the ship in port *within an agreed time*, known as the 'lay days'. The charter may agree to pay the shipowner at a certain rate for every day it improperly detains the ship beyond the agreed lay days. This is called 'demurrage'. It constitutes agreed damages for the charterer's breach of its obligation.[25] In other words, demurrage compensates the shipowner for the anticipated consequential loss arising from the charterer's delay through improper detention. The second point is the definition of 'deadfreight'. A charterer also undertakes a strict liability obligation to load *a full and complete cargo*. If the charterer fails to load the agreed quantity, it is liable to compensate the shipowner for the loss of freight equivalent to the cargo that was not loaded. This is called 'deadfreight'.

The line of decisions begins with the leading case of *Aktieselskabet Reidar v Arcos Ltd*.[26] Shipowners chartered a ship to charterers to load a full and complete cargo of timber by a certain date. The charterparty also provided for loading in an agreed time. The charterparty contained a demurrage clause requiring the charterers to pay liquidated compensatory damages at a fixed daily rate if they detained the ship in loading. The charterers exceeded the lay days.

---

[23] *J-Corp Pty Ltd v Mladenis* [2009] WASCA 157, (2010) 26 BCL 106 [8], [50].

[24] Stephen Furst and Vivian Ramsey, *Keating on Construction Contracts* (8th edn, Sweet and Maxwell 2006) [9-006] fn 50.

[25] *Suisse Atlantique Société d'Armement Maritime SA v NV Rotterdamsche Kolen Centrale* [1967] 1 AC 361 (HL) 395, 436.

[26] [1927] 1 KB 352 (CA).

If the charterers had completed loading the vessel within the time allowed, they would have loaded a summer cargo. However, by the time they loaded the ship, she could only legally carry a winter cargo of a much smaller quantity. The charterers, therefore, only loaded that smaller quantity. The English Court of Appeal held that the shipowners were entitled to recover demurrage at the fixed rate. However, the Court also held that the shipowners could recover general compensatory damages for the deadfreight, that is, the loss of freight caused by the charterers' failure to load a full and complete cargo. The general damages comprised the difference in freight between the summer cargo and the winter cargo.

Atkin LJ reasoned that 'The provisions as to demurrage quantify the damages, not for the complete breach, but only such damages as arise from the detention of the vessel.'[27] Given this statement, McGregor argues that a claimant may recover general compensatory damages in addition to liquidated compensatory damages 'if only part of the loss arising from a single breach is regarded as falling within the provision's ambit'.[28] Accordingly, the writer's interpretation suggests that a liquidated compensatory damages clause does not 'cover the field' regarding the compensatory response to a breach falling within its scope. The clause does not, therefore, provide an exclusive remedy.

Three appellate court judges decided the case. Atkin LJ formed part of the majority. However, properly read, the majority judgments do not support McGregor's interpretation of the case. Each separate judgment is considered below.

To start, Sargant LJ held that the charterers had breached two separate obligations, each giving rise to a distinct kind of loss.[29] The first was an obligation to complete loading the ship within the lay days. For this breach, the shipowners were entitled to the demurrage, being liquidated compensatory damages for the consequential loss flowing from the act of improperly detaining the vessel. The second was the obligation to load a full and complete cargo. For this breach, the shipowners were entitled to the deadfreight, being general compensatory damages for loss flowing from the act of failing to load the agreed quantity of cargo. The latter was a 'loss of another character'; 'a definite separate loss independent of … any loss arising from mere detention'.[30]

[27] *Aktieselskabet Reidar v Arcos Ltd* [1927] 1 KB 352 (CA) 363.

[28] Harvey McGregor, *McGregor on Damages* (19th edn, Sweet and Maxwell 2014) [15-024].

[29] These two duties are distinct: *Chandris v Isbrandtsen Moller Co Inc* [1951] 1 KB 240 (KB) 252.

[30] *Aktieselskabet Reidar v Arcos Ltd* [1927] 1 KB 352 (CA) 366-67.

Next, Bankes LJ held that the shipowner was able to recover general compensatory damages on a different basis.[31] The charterers had only breached a single obligation, to load the ship in the time allowed. However, this breach resulted in two 'essentially distinct' kinds of losses—one for detention of the vessel and one for loss of freight—and the clause only covered the former kind of loss. The problem with this analysis is that a charterer's act of detaining the vessel generally will not cause a loss of freight. The time at which the ship leaves port has nothing to do with the level of cargo it carries when it departs. Analogously, if a contractor is late in preforming building works for a principal, it has nothing to do with the quality of workmanship.

Lastly, as stated, Atkin LJ held that the demurrage clause quantifies 'the damages, not for the complete breach, but only such damages as arise from the detention of the vessel.'[32] Superficially, Atkin LJ appears to agree with Bankes LJ. However, that is not correct. The notion of an 'incomplete' breach is incoherent: a right is either violated, or it is not. The judgment is best interpreted as supporting the view that the charterer breached two separate obligations.[33] So, Atkin LJ held that the relevant breach for which the shipowners were entitled to recover general damages was the charterer's failure 'to *complete* the loading of the ship within the lay days'.[34] Consequently, the relevant 'breach' to which his Lordship referred was not a *mere* failure to load within the lay days, but a failure to load *a complete cargo* within that timeframe. Read in context, then, his Lordship's ambiguous statement is merely a shorthand method of referring to all of the charterers' breaches: namely, breach of the obligation to load a cargo within the stipulated time; and breach of the obligation to load a full and complete cargo. In short, like Sargant LJ, Atkin LJ was properly 'a "two breach" man'.[35] Further, that Atkin LJ saw the case as involving two independent kinds of loss is clear: he stated that the demurrage clause quantified only such loss as may have arisen 'from the detention of the vessel'.

Several cases have purported to apply the majority judgments in *Reidar*. One of them is Webster J's decision in *Total Transport Corp v Amoco Trading Co (The Altus)*.[36] The judge's reasoning is very difficult. On the dominant view, the shipowners in that case recovered

---

[31] *Aktieselskabet Reidar v Arcos Ltd* [1927] 1 KB 352 (CA) 358-59, 362.

[32] *Aktieselskabet Reidar v Arcos Ltd* [1927] 1 KB 352 (CA) 363.

[33] *Suisse Atlantique Société d'Armement Maritime SA v NV Rotterdamsche Kolen Centrale* [1965] 1 Lloyd's Rep 533 (CA) 539, 541 affd [1967] 1 AC 361 (HL) 389-90, 408, 418.

[34] *Aktieselskabet Reidar v Arcos Ltd* [1927] 1 KB 352 (CA) 363 (emphasis added).

[35] *The Bonde* [1991] 1 Lloyd's Rep 136 (Com Ct) 141.

[36] [1985] 1 Lloyd's Rep 423 (Com Ct).

general compensatory damages *and* liquidated compensatory damages for the charterer's breach of the same duty—to load a full and complete cargo.[37] For three reasons, however, the decision is flawed. First, Webster J interpreted Atkin LJ in *Reidar* as holding that the charterer breached only one obligation, not two.[38] That was a mistake. Second, the judge failed to consider the relevant decision of *Suisse Atlantique Société d'Armement Maritime SA v NV Rotterdamsche Kolen Centrale*.[39] Finally, in *Richco International Ltd v Alfred C Toepfer International GmbH (The Bonde),* Potter J closely analysed the judgments in *Reidar*, and several judgments interpreting that case. In the end, Potter J declined to follow *The Altus*.[40] The English High Court has since applied the decision in *The Bonde*.[41] One must conclude, then, that *The Aldus* is no longer good law.

In conclusion, properly explained *Reidar* is a case where, despite the fact a liquidated compensatory damages clause existed, A recovered from B general compensatory damages for an independent kind of loss flowing from B's breach of a separate contractual duty owed to A. This allegedly 'difficult' case merely illustrates that any exclusivity only applies for the particular breach or breaches, and kinds of loss, falling within the provision's scope. Several cases have sought to follow *Reidar*. Some of them have misinterpreted the majority judgments of Sargant LJ and Atkin LJ. Those cases are no longer good law.

III Principled arguments

As best interpreted, Anglo-Australian law unequivocally supports that view that liquidated compensatory damages are an exhaustive remedy. Is this view the most principled?

Rossiter suggests it is not. The writer argues that, in principle, liquidated compensatory damages clauses should provide a non-exclusive remedy, reasoning that 'in the normal course, one would expect that an agreed damages clause is inserted … solely for the benefit of [A]'.[42] The writer's reasoning is flawed. Generally, an agreed compensatory damages clause is

---

[37] *The Bonde* [1991] 1 Lloyd's Rep 136 (Com Ct) 141-42; Harvey McGregor, *McGregor on Damages* (19th edn, Sweet and Maxwell 2014) [15-024].

[38] *The Altus* [1985] 1 Lloyd's Rep 423 (Com Ct) 434-35.

[39] [1965] 1 Lloyd's Rep 533 (CA), affd [1967] 1 AC 361 (HL).

[40] [1991] 1 Lloyd's Rep 136 (Com Ct) 141-42.

[41] *ERG Raffinerie Mediterranee SpA v Chevron USA Inc (The Luxmar)* [2006] EWHC 1322 (Com Ct), [2006] 2 All ER (Comm) 913 [58], affd [2007] EWCA Civ 494, [2007] 2 All ER (Comm) 548.

[42] Chris Rossiter, *Penalties and Forfeiture* (Lawbook Company 1992) 71.

inserted for *both* parties' benefit.[43] Consequently, there is no good reason to permit A to recover more than the liquidated compensatory damages amount.

In fact, for several reasons, the law's preference for exclusivity is principled. Three arguments follow.

First, assume that liquidated compensatory damages clauses provided a non-exhaustive remedy. It follows that, if A incurred losses beyond the liquidated compensatory damages sum, she should be able to claim general compensatory damages for those losses. If that were true, one would expect mutuality. Accordingly, if A incurred losses less than the liquidated compensatory damages amount, B should be able to prove this fact, and reduce his exposure. But no one supports this move.

Second, if liquidated damages clauses do not provide an exclusive remedy, the main advantages of these clauses are eliminated.[44] For example, each party cannot know in advance what A's right, and B's duty, will be upon B's breach. No certainty will exist.

Finally, a central feature of an agreed damages clause is that it supplants a court's jurisdiction to determine A's general damages. Effectively, the parties 'step into the shoes' of the court for the purpose of awarding the damages payable for B's breach. It follows that A cannot claim from B general damages in addition to, or instead of, liquidated damages because when the parties include an agreed damages clause in their contract, they 'contract out' of the general law damages regime. The agreed damages regime stands in its place. Thus, in *Chattan Developments Ltd v Reigill Civil Engineering Contractors Ltd*, Ramsey J held that liquidated damages are the 'sole remedy' for any particular breach to which a liquidated damages clause relates; and that general damages 'are not recoverable because the parties' agreement of liquidated damages *replaces* the remedy which would otherwise be available for breach'.[45] The concept of 'replacement' refers to the 'contracting out' that takes place. However, the 'contracting out' is limited to breaches falling within the clause's scope.

2 Penalty clauses

A Nature

[43] *Suisse Atlantique Société d'Armement Maritime SA v NV Rotterdamsche Kolen Centrale* [1967] 1 AC 361 (HL) 420-21.

[44] Text accompanying and following n 16 in ch 1.

[45] [2007] EWHC 305 (TCC) [31] (emphasis added). See similarly *Carter on Contract* [33-370] (service 37, April 2014).

This section deals with two questions regarding the nature of penalty clauses. First, properly comprehended, in the paradigm scenario[46] are penalties in the nature of general damages, or a debt? Second, are penalty clauses correctly unenforceable, or void?

As to the first question, Chapter 3 illustrated that historically A was able partially to enforce a penalty clause. Thus, traditionally the English Courts viewed the enforceable part of the penalty as being in the nature of a debt. For example, in *Lowe v Peers*, Lord Mansfield held that in the case of 'covenants secured by a penalty', A was entitled to 'bring an action of debt for the penalty'.[47] Later, in *Beckham v Drake*, Lord Campbell held that the recoverable portion of a penalty is 'in the nature of a debt', such that 'an action of debt might be maintained for it'.[48] In contemporary times, the English Court of Appeal[49] and the Australian High Court[50] have held that an action on a penalty clause is an action in debt. Thus, Andrews notes that, as a matter of doctrine, a claim on a penalty clause is 'technically a debt claim, rather than one for damages'.[51]

In principle, penalties display some of the indicia of a debt, but not others. On the one hand, in some cases, A may seek summary judgment against B based upon the penalty clause.[52] Further, in principle, usually A will become an unsecured creditor upon B's insolvency. On the other hand, A must prove that she incurred loss in order to recover, as she only recovers to the extent of such proof. Further, as A must prove her loss, in some cases her penalty claim will be unliquidated. The interaction between penalty clauses and the remoteness and mitigation doctrines is difficult; it will be considered in a moment. At its highest, then, it might be said that penalties generally are *akin* to a debt.

As to the second question, the terms 'unenforceable' and 'void' permeate Anglo-Australian private law discourse. A 'void' contract produces 'no legal effects whatsoever'.[53] It is for all purposes disregarded, and treated as non-existent.[54] Thus, a void contract does not give rise to legally-*binding* obligations. Conversely, 'unenforceable' contracts come into

---

[46] Text following n 20 in ch 1.

[47] (1768) 4 Burr 2225, 2228; 98 ER 160, 162.

[48] (1849) 2 HL Cas 579, 645; 9 ER 1213, 1237.

[49] *Jobson v Johnson* [1989] 1 WLR 1026 (CA) 1040-41.

[50] *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 202, 204, 212.

[51] Neil Andrews, *Contract Law* (CUP 2011) [19.19] (footnote omitted).

[52] *Gerrard v Clowes* [1892] 2 QB 11 (QB) 13.

[53] Hugh Beale (gen ed), *Chitty on Contracts* (2nd supp, 31st edn, Sweet and Maxwell 2013) [1-100].

[54] *Boddington v Lawton* [1994] ICR 478 (Ch) 491.

existence, however one party, or both, cannot successfully sue on the contract.[55] Hence, an unenforceable contract does not give rise to legally-*enforceable* obligations.

A penalty clause is either unenforceable or void. Until recently, English and Australian law seemingly diverged. We consider English law first, then Australian law.

In England, the weight of judicial opinion has always favoured the view that penalty clauses are partially unenforceable. Three appellate court examples will suffice. First, in *Beckham*, Maule J held that a penalty clause is not 'rendered wholly inoperative, but it retains the effect of binding the failing party to pay such part of the sum as may be reasonable in respect of the failure …'.[56] Second, in *Jobson v Johnson*, Nicholls LJ held that a penalty clause 'remains in the contract and can be sued upon, but it will not be enforced by the court beyond the sum' representing A's incurred loss.[57] Similarly, Kerr LJ held that, if a clause is penal, 'to that extent it is unenforceable. But it is enforceable to the extent that it is not a penalty …'.[58] Finally, in *Makdessi v Cavendish Square Holdings BV*, the English Court of Appeal approved *Jobson,* and expressly confirmed that penalty clauses are unenforceable.[59]

One 'divisive' case exists. In *Azimut-Benetti SpA (Benetti Division) v Healey*,[60] Blair J held that if a contract between A, a creditor, and B, a principal debtor, contains a penalty clause, A cannot enforce the clause against C, the guarantor of B's obligations. The judge gave two reasons. For one, this conclusion was said to follow from the fact that the Penalty Rule is based on public policy. For another, it was held that a penalty clause does not give rise to any obligation in B, as the clause is void. It followed that B can be under no duty the performance of which C can guarantee. This reasoning is wrong. The first reason purports to rely on illegitimate policy-based reasoning. Further, although Blair J attempted to justify his decision in public policy terms, the judge did not identify the exact 'policy' that penalty clauses offend. Accordingly, even if policy-based reasoning were legitimate, the proposed 'justification' is entirely empty. The second reason is flatly inconsistent with *Jobson*, a case not brought to the judge's attention.

---

[55] Hugh Beale (gen ed), *Chitty on Contracts* (2nd supp, 31st edn, Sweet and Maxwell 2013) [1-104].

[56] *Beckham v Drake* (1849) 2 HL Cas 579, 622; 9 ER 1213, 1229.

[57] [1989] 1 WLR 1026 (CA) 1040.

[58] *Jobson v Johnson* [1989] 1 WLR 1026 (CA) 1049. See also 1035 (Dillon LJ).

[59] [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [1], [47]-[53], [125].

[60] [2010] EWHC 2234 (Com Ct), [2011] 1 Lloyd's Rep 473 [24].

Conversely, for many years, a penumbra of uncertainty plagued the Australian private law landscape. Today, the horizon is clearer. The story has unfolded in three chapters.

To start, modern appellate authority first viewed penalty clauses as unenforceable. For example, in *IAC (Leasing) Ltd v Humphrey*, Walsh J held that the Penalty Rule rendered an agreed damages clause unenforceable.[61] Similarly, in *O'Dea v Allstates Leasing System (WA) Pty Ltd*, Murphy J, Wilson J and Deane J described a penalty clause as unenforceable.[62]

Next, the Australian High Court's decision in *AMEV-UDC Finance Ltd v Austin* created uncertainty: on the one hand, Mason and Wilson JJ held that a penalty is 'unenforceable, or perhaps void, *ab initio*'; on the other hand, in dissent Deane J held that 'a penal clause is not void'; rather, it is unenforceable 'to the extent to which liability under the clause exceeds the true damnification'.[63]

Superficially, following *AMEV,* appellate court decisions favoured unenforceability. For example, in *Citicorp Australia Ltd v Hendry*, Priestley JA held that a penalty clause is unenforceable, meaning that 'it has no legal effect; the party for whose benefit it would operate if it was enforceable can at no stage enforce it or obtain the help of the law in any way in deriving any benefit from it. It is the same as if it was not in the agreement at all.'[64] However, although these appellate court judges adopted the label 'unenforceable', the clause they described possessed all the characteristics of a void clause. The 'unenforceable' penalty clause they described produced 'no legal effects whatsoever'; was for all purposes disregarded; and was treated as non-existent. As Diplock LJ once held, if a contract is found 'not to be a contract at all', then despite being labelled 'unenforceable', it may 'accurately be said to be void'.[65] This statement similarly applies to Priestley JA's concept of a penalty clause. In principle, this must be correct. If what his Honour described is in truth unenforceability, there is no real difference with voidness, rendering any distinction meaningless. At least one appellate Australian court saw fit to call a spade a spade, declaring that a penalty clause was, at the time, void.[66]

---

[61] (1972) 126 CLR 131 (HCA) 143.

[62] (1983) 152 CLR 359 (HCA) 375, 378, 397, 401 cf 372.

[63] (1986) 162 CLR 170 (HCA) 192, 203.

[64] (1985) 4 NSWLR 1 (NSWCA) 39.

[65] *Mackender v Feldia AG* [1967] 2 QB 590 (CA) 601.

[66] *Ronstan International v Thomson* [2002] VSCA 75 [25].

Lastly, the Australian High Court's decision in *Andrews v Australia and New Zealand Banking Group Ltd* reinstated the original view.[67] The Court rejected the then 'conventional view'[68] that a penalty clause was void. Rather, the High Court held that, in a case involving an invalid agreed compensatory damages clause, A is permitted to enforce the clause to the extent of her incurred loss, and B is 'relieved to that degree from liability'.[69]

Worthington argues that, as a matter of *law*, penalty clauses are 'unenforceable except to the extent that they represent [A's] real loss'.[70] Rossiter suggests that, as a matter of *principle*, partial unenforceability presents the 'better view'.[71] As it stands today, then, Anglo-Australian private law is principled.

At this point, it is important to note precisely how a penalty clause works. To do so, we must correctly understand the nature of the rights here in play.

In Chapter 1, this thesis outlined that the term 'right' was ambiguous, and that it in fact concealed four distinct types of legal entitlement: a claim, a privilege, a power and an immunity. It further noted that where a party holds one of these four types of 'rights', a counterparty occupies a correlative, specific normative position: a duty, 'no-right', a liability and a disability respectively. For present purposes, two meanings of the word 'right' are crucial. The first is a 'claim right', which correlates with a legal duty. The second is a 'power', which correlates with a liability. Scholars refer to this power as a 'power to sue' or a 'right of action'. As Nolan and Robertson explain, 'Where a private law right is violated, this will usually trigger a right of action, which in Hohfeldian terms is a power in the person wronged to bring an action against the wrongdoer. This power correlates with a "liability" of the wrongdoer, in the sense that the wrongdoer is now susceptible to legal action.'[72] The extent to which a defendant is subject to the claimant's power to sue for damages will determine the degree of his liability. Various things may limit the extent of a defendant's liability, such as a limitation period expiring; the parties failing to put a contract in a required form; or a limitation or exclusion clause.

---

[67] [2012] HCA 30, (2012) 247 CLR 205.

[68] JW Carter and others, 'Contractual Penalties: Resurrecting the Equitable Jurisdiction' (2013) 30 JCL 99, 114.

[69] *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [10].

[70] Sarah Worthington, *Equity* (2nd edn, OUP 2006) 224.

[71] Chris Rossiter, 'Relief Against Penalties' in Patrick Parkinson (ed), *Principles of Equity* (2nd edn, Lawbook Company 2002) [801] fn 1.

[72] Donal Nolan and Andrew Robertson, 'Rights and Private Law' in Donal Nolan and Andrew Robertson (eds), *Rights and Private Law* (Hart Publishing 2012) 15-16 (footnote omitted).

Where a court holds that a clause is properly penal, the correct analysis is this. A's *claim right* to the entire penal sum remains in place, although it is partially unenforceable; concomitantly, B's *obligation* to pay the entire penal sum remains in place, although it is only partially enforceable. However, A's *right of action* to recover (or power to sue for) the entire penal sum is truncated, such that she is only permitted to enforce the claim right to the extent of her incurred loss; concurrently, B's *liability* to satisfy the entire penal sum is cut down to an amount equivalent to A's incurred loss. Properly understood, then, a penalty clause works in a similar way to a limitation clause, in that it is the claimant's power to sue—but not her claim right—that is truncated. A significant difference, of course, is that typically a penalty clause contains a very high ceiling, whereas a limitation clause contains a low one.

The High Court's approach in *Andrews*, and the Penalty Rule's history, supports this analysis. As to the former, when describing a penalty clause's effect, the Court described A's *right* to a penalty as being enforceable, but only to a limited extent, and it also described B as being relieved from *liability*. It follows inferentially that the Court saw both concepts—the partial enforceability of B's obligation *and* the cutting down of B's liability—as relevant to a penalty clause analysis. As to the latter, recall that historically if a plaintiff instituted proceedings to recover a penal sum due on a bond, courts gave judgment for the full amount.[73] By doing so, they essentially acknowledged the plaintiff's claim right to the entire penal sum. However, courts would only permit the plaintiff to recover for her incurred loss. Effectively, they truncated the plaintiff's right of action to accord with her sustained loss.

B Incurred loss *less* than penalty

Assume that A's incurred loss is *less* than the penalty amount. If so, what is the means by which A may recover the amount of her incurred loss? Further, does this situation raise any problematic issues? If it does, are the problems solvable? This section addresses those questions. It begins by analysing the apparent dual means of recovery. It then evaluates three theoretical problems that have beset this area of the law.

I Means of recovery

As Chapter 3 outlined, orthodoxy holds that if the court finds an agreed compensatory damages is invalid, and A's incurred loss is less than the penalty amount, one of two consequences may follow. On the one hand, A may sue B for general compensatory damages

---

[73] During the 18th century, the general law and legislative approaches were broadly the same: see text to nn 24-25 in ch 2.

for breach of contract. On the other hand, the clause may be 'compressed' to reflect A's incurred loss, and A may sue B on the penalty clause to recover the amount of that loss. Accordingly, two potential means of recovery seemingly exist.

As Chapter 2 discussed, historically A maintained a right of election. She could choose to sue on the penalty clause in debt, or claim general damages in assumpsit.[74]

The tradition has seemingly continued in England, as modern cases exist which support each of these options. For example, in *Robophone Facilities Ltd v Blank*, Diplock LJ held that if a clause is held penal, A is relegated to her right to claim from B general compensatory damages for breach of contract.[75] Conversely, in *Jobson*, Nicholls LJ held that, as the penalty clause remains in the contract, A may choose to sue on it, but the penalty will not be enforced beyond the amount of A's incurred loss.[76] His Lordship held that, if A claims on the penalty clause, the court will engage in a 'scaling down' exercise: it will relieve by 'cutting down the extent to which the contractual obligation is enforceable'.[77] The court scales down the penal amount to a sum that equals A's incurred loss. Put differently, the clause is 'compressed' to reflect A's incurred loss, and A may then sue B upon the clause to recover. Recently, in *Makdessi*, the English Court of Appeal endorsed Nicholls LJ's reasoning in *Jobson*.[78]

Contrast the Australian position. In *AMEV*, the majority of the Australian High Court held that A maintained no right to elect: she could only sue for general damages at law. So, Mason and Wilson JJ held that 'Without exception the cases show that once the agreed sum is held to be a penalty the [claimant] recovers damages for breach of contract in lieu of the penalty.'[79] A penalty clause, it was said, could not be partially enforced. Conversely, in dissent Deane J held that A is entitled partially to enforce a penalty clause up to the amount of A's actual incurred loss.[80]

Following *AMEV*, the English Court of Appeal decided *Jobson*. As a result, this caused some Australian judges to question whether A was indeed only entitled to claim against

---

[74] *Beckham v Drake* (1849) 2 HL Cas 579, 615, 645; 9 ER 1213, 1226, 1237.

[75] [1966] 1 WLR 1428 (CA) 1446.

[76] *Jobson v Johnson* [1989] 1 WLR 1026 (CA) 1040.

[77] *Jobson v Johnson* [1989] 1 WLR 1026 (CA) 1042.

[78] *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [125]-[126].

[79] *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 191. See similarly 175, 176 (Gibbs CJ).

[80] *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 203, 205.

B general damages where a penalty clause existed. It was said that Australian private law awaited the proper occasion to review that principle.[81]

The occasion came in *Andrews*. As noted, the High Court held that A is entitled to enforce a penalty clause to the extent of her incurred loss, and B is relieved to any further degree from liability.[82] Also, the Court endorsed the 'scaling down' exercise to which Nicholls LJ referred in *Jobson*.[83] These statements are only consistent with A maintaining a truncated right of action against B based on the penalty clause. In *Paciocco v Australia and New Zealand Banking Group Ltd*, Gordon J confirmed this interpretation.[84] Her Honour took Deane J's dissenting view in *AMEV* to be consistent with the Court in *Andrews*, and hence to represent the law in Australia. It followed that, on the one hand, penalty clauses are partially unenforceable, but 'only to the extent that the amount stipulated … exceeded the quantum of the relevant loss'; and, on the other hand, such clauses are partially enforceable, such that A may 'enforce the stipulation that is found to be a penalty to the extent of [her] loss …'.[85] Further, in *Cedar Meats (Aust) Pty Ltd v Five Star Lamb Pty Ltd*, the Victorian Court of Appeal agreed with her Honour's analysis.[86]

Carter and others argue that it is incorrect to allow A to sue upon the penalty clause, and partially enforce it, for three reasons.[87]

First, the writers argue that, in allowing partial enforcement, the High Court's decision in *Andrews* contravenes the majority decision in *AMEV*. However, when one considers the body of Anglo-Australian cases as a whole, it the latter judgments that appear to jar.

Second, Carter and others propound that, although the High Court put forward A's power to sue on the penalty clause as a general rule, 'there are no modern cases to support it in the context of promises to pay money'.[88] However, the Court of Appeal's decision in *Jobson* provides modern case law support in the context of an obligation to transfer a proprietary right. In the review jurisdiction context, the subject matter of the obligation does not form the

---

81 *PC Developments Pty Ltd v Revell* (1991) 22 NSWLR 615 (NSWCA) 645.

82 *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [10].

83 *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [60]-[61].

84 [2014] FCA 35, (2014) 309 ALR 249.

85 *Paciocco v Australia and New Zealand Banking Group Ltd* [2014] FCA 35, (2014) 309 ALR 249 [48].

86 [2014] VSCA 32 [56].

87 JW Carter and others, 'Contractual Penalties: Resurrecting the Equitable Jurisdiction' (2013) 30 JCL 99, 115, 117.

88 JW Carter and others, 'Contractual Penalties: Resurrecting the Equitable Jurisdiction' (2013) 30 JCL 99, 115.

basis for any principled distinction.[89] Why should it make a difference regarding consequential issues?

Finally, these scholars contend that it is 'peculiar' that A should have a choice of remedies if the provision is penal, but only one remedy where the clause is liquidated damages. True enough. It has been held that a penalty clause is 'in practice, a dead letter',[90] because A's recovery under the clause and at general law is substantially equal.[91] It is said to follow that A's right of action on the penalty disappeared from the law, leaving her only with a right to claim general damages. But if two methods of recovery existed, why did A's right to general compensatory damages trump her power to sue on the penalty clause? It can equally be said that, given the same quantum of recovery, the cases are best interpreted as ones where A was enforcing the clause. In fact, this interpretation is preferable for at least two reasons.

For one, this interpretation is more transparent. It adheres more to the courts' general reasoning in the appellate court decisions in *Jobson, Makdessi* and *Andrews*.

For another, this stance is more theoretically sound.

On the one hand, as discussed, the fact that the parties included an agreed damages clause in their contract means they 'contracted out' of the general law damages regime. This supplanted the court's power to award general damages for breach. If a penalty clause properly is void, then it is for all purposes disregarded, and treated as though it never existed. Consequently, the court's jurisdiction to determine A's general damages is reinstated. However, correctly understood, a penalty clause is partially enforceable, and hence remains in the contract. Accordingly, the court's jurisdiction remains ousted; the consensual damages regime continues to apply in place of the general law damages regime. Therefore, A cannot bring a general damages claim.

On the other hand, a clause may provide that B will incur a detriment upon a substantive breach. Chapter 4 argued that an Anglo-Australian court might properly classify such a clause as penal. If so, A will maintain no right to claim general compensatory damages to recover her incurred loss, as there is no actual breach upon which to base her claim. Conversely, A will maintain a power to sue B on the penalty clause. A may partially enforce the clause to recover her incurred loss, as an actual breach is unnecessary for this claim.

---

[89] Text to n 4 and n 8 in ch 4.

[90] *Jobson v Johnson* [1989] 1 WLR 1026 (CA) 1039. See similarly *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 212.

[91] Harvey McGregor, *McGregor on Damages* (19th edn, Sweet and Maxwell 2014) [15-026]. See also *Jobson v Johnson* [1989] 1 WLR 1026 (CA) 1040.

II Problematic issues

(1) The assessment date

This thesis has argued that, if a court invalidates an agreed compensatory damages clause, A may sue B on the penalty clause and enforce it up to a sum equivalent to her incurred loss, assessed at the judgment date. However, the test of validity is applied at the contract date. Given this, if the agreed compensatory damages clause is penal, why is A not permitted to recover from B a suitable amount, also evaluated *at the contract date*?

Consider the following example. Suppose that, in the paradigm scenario,[92] the agreed compensatory damages clause states that B has a duty to pay A £1,000 if B breaches the contract. A court considers that the clause fails the test of validity, and is penal. As at the contract date, A's anticipated losses were £600. As at the judgment date, A's incurred losses total £500. The court must then conduct a 'scaling down' exercise, 'cutting down' to the extent to which the contractual obligation is enforceable. It will scale down the amount to £500; the sum that equals A's *incurred* loss, assessed *ex post*. Why is A not permitted to recover £600—her *anticipated* loss—given that the court determined the clause's validity *ex ante*?

The answer to this problem is that A recovers her incurred loss, assessed *ex post*, because that is the parties' implied intentions. So, when the parties agree to an agreed damages clause, they do so seeking to obtain the certainty that a valid agreed damages clause will provide. The parties wish to be reasonably certain of the extent of each party's rights and obligations, and 'the risks which they run as a result of entering into the contract'.[93] The certainty in play is an *effect* of a liquidated damages clause.[94] If a dispute ensues, the only way to achieve the desired effect of certainty is to have the court evaluate the clause's validity as at the contract date, as it likely will result in a court upholding the clause. Accordingly, the parties implicitly agree that a court will assess the validity of the agreed damages clause *ex ante*.

Two points are of note. The first is that this agreement is not express; rather, it forms a background assumption to the parties' interactions. This is unproblematic, for contracting parties can intend things they do not mention, or even think about, during negotiations.[95] The

---

[92] Text following n 20 in ch 1.

[93] *Philips Hong Kong Ltd v A-G of Hong Kong* (1993) 61 BLR 41 (PC) 54.

[94] If the *purpose* of all agreed damages clauses were to achieve certainty, there would be no Penalty Rule at all: the private law would simply enforce all agreed damages clauses.

[95] See Charles Fried, *Contract As Promise: A Theory of Contractual Obligation* (Harvard UP 1981) 88.

second is that the desire to achieve certainty provides the motive for the parties' agreeing to an *ex ante* assessment date.

If the clause is upheld, the parties obtain the benefit of the certainty that they seek. Conversely, if the clause is invalidated, certainty is no longer in issue as, by definition, the court must revise the amount A will recover. Accordingly, there is no reason for the parties to agree to an *ex ante* assessment date in these circumstances. Put differently, as the sought-after effect of certainty which motivated the original agreement as to the assessment date can no longer exist, there is no sense in the parties impliedly agreeing to have the court assess A's loss as at the contract date. As such, necessarily the parties must implicitly agree to an *ex post* evaluation where the clause is held penal.

Importantly, the parties' implied agreement is principled. This is because evaluating the revised amount under a penalty clause as at the contract date would leave the parties with the worst of both worlds. A liquidated damages clause sacrifices exactitude of assessment because it provides valuable, *timely* certainty (that is, the clause provides a sum certain at a time when parties can plan their future affairs upon that knowledge). Where A's loss is assessed *ex post*, a penalty clause provides exactitude of assessment at a time when certainty is unimportant: the date at which the parties required certainty (to allow for planning) has long passed. Conversely, if A's loss is assessed *ex ante*, a penalty clause would sacrifice exactitude of assessment, whilst failing to provide timely certainty.

(2) The interaction with remoteness and mitigation

This thesis has contended that, in the paradigm scenario,[96] a penalty is akin to a debt. As previous mentioned, two indicia of a debt are that the remoteness doctrine is irrelevant, and the mitigation doctrine does not generally apply. However, in principle, if A sued partially to enforce a penalty clause, she would be required to prove her loss, such that the claim would be subject to the remoteness and mitigation doctrines.[97] A dissonance seemingly exists.

This discord will be more theoretical than real. In principle, the legal doctrines limiting general compensatory damages, such as remoteness and mitigation, will apply to A's claim to recover under a penalty clause. But, in fact, in most cases those doctrines will apply in such a way that they will not affect A's quantum of recovery.

On the one hand, the remoteness doctrine will not likely restrict the amount of A's

[96] Text following n 20 in ch 1.

[97] *GWC Property Group Pty Ltd v Higginson* [2014] QSC 264 [43], [46]-[47].

recovery. By the agreed damages clause, B expressly undertook responsibility for A's losses of a particular type, *up to and including* the agreed compensatory damages amount. That amount is then held penal. Per hypothesis, A's incurred loss is less than the penalty amount. Accordingly, whatever A's incurred loss, B will have assumed responsibility to compensate A for it, and it cannot be too remote. So, suppose an agreed compensatory damages clause obliges B to pay to A £1,000 if B breaches the contract. A court invalidates the clause. Whether A's incurred loss is £500 or £800, B will have undertaken responsibility to compensate A for the relevant loss. It will not be too remote.

On the other hand, assume that A's anticipated losses and anticipated mitigation opportunities are difficult to ascertain, and the clause is held penal. The mitigation doctrine does not require a claimant to act in accordance with best practice, or even with the strict law: the doctrine only requires her to act *reasonably*.[98] So, will it be unreasonable for A not to have mitigated her loss when the proper nature of the agreed compensatory damages clause is unclear; courts are admittedly 'predisposed' to finding that such a clause is valid;[99] and the mitigation doctrine does not affect the clause if valid? In most cases, the answer is probably 'no'.

In conclusion, if A sues B on a penalty clause, in most cases the remoteness and mitigation doctrines will not reduce the amount she recovers. Therefore, a penalty clause claim is usually similar to a debt claim.

(3) The relevance of unjust enrichment

In *Andrews*, the applicants claimed that the clauses requiring bank customers to pay various 'exception fees' were penal. However, when the claim was brought, the applicants had already paid the fees to the respondent bank. The claim, therefore, was to be *repaid* the fees. Accordingly, the Australian High Court described the claim as being for moneys had and received by the bank to its use.[100] An action for 'money had and received to the use of the plaintiff' is merely an archaic label for an unjust enrichment claim.[101] So, in modern terms, the applicants claimed restitution of the fees in an action in unjust enrichment.

---

[98] Edwin Peel, *The Law of Contract* (13th edn, Sweet and Maxwell 2011) [21-013].

[99] *Alfred McAlpine Capital Projects Ltd v Tilebox Ltd* [2005] EWHC 281 (TCC), [2005] BLR 271 [48(3)]; *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [75(iii)].

[100] *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [6].

[101] *Fibrosa Spolka Akcyjna v Fairbairn Lawson Combe Barbour* [1943] AC 32 (HL) 61-63, 65; *Baltic Shipping Company v Dillon (The Mikhail Lermontov)* (1993) 176 CLR 344 (HCA) 355.

Carter and others find the Court's analysis troubling. On the one hand, the writers argue that the Penalty Rule and a claim for restitution in unjust enrichment are strangers, as the nature of the consequential relief provided under the rule 'counts against any claim being made for restitution'.[102] On the other hand, Carter and others assert that if penalty clauses are properly partially unenforceable, and B has already paid the penalty amount, any claim by B for restitution in unjust enrichment is irrelevant. Conversely, if penalty clauses are properly void, a basis for restitution would exist.[103]

These writers' concerns are misplaced. In truth, where B has already satisfied his obligation under a penalty clause, commonly an unjust enrichment claim will not only be relevant, but essential to do justice as between the parties. This is so regardless of whether a penalty clause is properly partially unenforceable, or void. These statements need unpacking. Three points are necessary. For simplicity, the analysis is limited to a case where, in the paradigm scenario,[104] B has made payment to A under the penalty clause before proceedings are instituted.

To start, traditionally courts have seen two approaches open when a clause is penal. These approaches can be described as the 'expunction' approach and the 'compression' approach. Under the former view, a penalty clause is void. The clause is 'expunged' from the contract, and the agreed damages regime is a nullity. With the general damages regime reinstated, A is able to bring a general damages claim against B to recover her incurred loss. Under the latter view, a penalty clause is partially unenforceable. The agreed damages regime subsists, although the clause is 'compressed' to the extent of its enforceability, which means B's liability is 'cut down' to an amount equivalent to A's incurred loss. A is then able to bring a claim partially to enforce the penalty clause to that extent.

Next, if one takes the expunction approach, the relevant clause is properly void. Effectively, the court will treat the clause as though it never existed. A will have a claim right against B for breach of contract. She will claim general compensatory damages for her incurred loss. However, B will hold a counterclaim right in unjust enrichment against A, usually based on the unjust factor of mistake.[105] The mistake is one of law: namely, a mistake as to the legal *existence* of the clause. B will claim restitution of the penalty amount transferred.

---

[102] JW Carter and others, 'Contractual Penalties: Resurrecting the Equitable Jurisdiction' (2013) 30 JCL 99, 108.

[103] JW Carter and others, 'Contractual Penalties: Resurrecting the Equitable Jurisdiction' (2013) 30 JCL 99, 108.

[104] Text following n 20 in ch 1.

[105] In some cases, the relevant unjust factor may be failure of basis.

In the premise, A's incurred loss is less than the penalty amount. It follows that once B's restitutionary right is set off against A's compensatory right, A will be required to pay to B the amount by which the penalty sum exceeds A's incurred loss. For example, assume that a contract contains an agreed compensatory damages clause requiring B to pay to A the sum of £1,000 if B breaches the contract. B breaches his duty, and then pays. A incurs losses of £100. The court holds the clause penal, and expunges it. B's unjust enrichment claim will total £1,000.[106] A's breach of contract claim against B will total £100. After set off, B will recover £900 from A.

Lastly, Anglo-Australian private law has settled on the compression approach. Although this approach provides for a different theoretical analysis, an unjust enrichment claim remains relevant, and the outcome is the same. On this view, the clause is partially unenforceable. The court will compress the clause, 'scaling down'[107] the penalty amount to an amount equal to A's incurred loss. A will have a right of action against B to enforce the penalty clause to that extent. Accordingly, A will be entitled to retain an amount equal to that scaled down sum. However, B will hold a counterclaim right in unjust enrichment against A based on mistake. As with the previous scenario, B will claim restitution. The mistake is again one of law, but of a different nature: namely, it is a mistake as to the legal *enforceability* of part of the clause. Accordingly, B is only entitled to recover from A to the extent of his mistake, that being an amount equal to the unenforceable part of the clause. So, as regards the sum representing the unenforceable part of the clause, A is not entitled to retain it; and B is entitled to recover it. In short, B has an unjust enrichment claim right to restitution of a sum reflecting the unenforceable component of the clause—that is, the excessive penal element. Given A's incurred loss is less than the penalty amount, A will be required to pay to B the difference in amount between the penal sum and A's incurred loss. So, consider again the example given. B pays to A the £1,000 penalty sum. This time, the clause is partially unenforceable. Consequently, the penal sum is scaled down to a sum equal to A's incurred loss of £100. A's penalty clause right of action is worth £100; B's unjust enrichment claim right is worth £900.[108] No set off is required. Again, B will recover £900 from A.

---

106 Assuming that A cannot raise a valid defence.

107 *Jobson v Johnson* [1989] 1 WLR 1026 (CA) 1042.

108 Again, assuming no defences exist.

In *Andrews*, ultimately the High Court remitted back the case to the Federal Court to decide. It did so in *Paciocco*. Gordon J's judgment supports this analysis.[109] The judge held that the applicants had a claim right in unjust enrichment for restitution based on a mistake of law. Further, Gordon J held that the value of the unjust enrichment claim right was calculated by reference to the difference between the penal late payment fee and the respondent bank's loss consequent upon the relevant customer's late payment.[110]

Unjust enrichment, then, is not irrelevant to the consequential relief the Penalty Rule provides. In fact, it is essential to a proper working out of the parties' rights in the agreed damages context. Unjust enrichment explains why, as Birks once noted, 'since a promised penalty cannot be claimed, a penalty which has already passed into the other's hands [is] recoverable'.[111]

C Incurred loss *more* than penalty

In rare situations, A's incurred loss may in fact *exceed* the penalty amount. This seemingly paradoxical situation might arise, for example, where a moderate agreed compensatory damages amount nevertheless fails the Estimation Test, because the sum was not 'substantially accurate'.

If A is permitted partially to enforce a penal clause, logically she cannot recover more than the penal sum.[112] Nothing more needs to be, or can be, said.

Conversely, if A may sue B for breach of contract, then the following 'controversial issue' arises: where A's incurred loss is greater than the penalty amount, can A disregard the penalty clause and recover her incurred loss?[113] Put differently, can A claim against B general compensatory damages for her entire legally-recoverable loss without limit, or does the penal sum 'cap' her recovery? The cases in this area cannot be reconciled. As between the two

---

[109] *Paciocco v Australia and New Zealand Banking Group Ltd* [2014] FCA 35, (2014) 309 ALR 249 [359], [363], [365]-[366].

[110] This analysis does not fall foul of the 'defendant's legal entitlement' qualification to an unjust enrichment claim: Andrew Burrows, *The Law of Restitution* (3rd edn, OUP 2011) 88-91. One recognised exception to that qualification is where the claimant's contractual obligation is unenforceable for want of formality: *Deglman v Guaranty Trust Co of Canada* [1954] SCR 725 (SCC); *Pavey and Matthews Pty Ltd v Paul* (1987) 162 CLR 221 (HCA). In principle, the reason why the obligation is unenforceable does not provide a valid basis for distinction. It follows that B's unjust enrichment claim is relevantly exceptional, and should succeed despite the fact that B is subject to an extant obligation.

[111] Peter Birks, *An Introduction to the Law of Restitution* (Clarendon Press 1985) 215.

[112] *Wall v Rederiaktiebolaget Luggude* [1915] 3 KB 66 (KB) 72-73; *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 202.

[113] Andrew Burrows (ed), *English Private Law* (3rd edn, OUP 2013) [21.180].

options, the preferable answer is difficult to discern. Leading scholars sharply divide on the point.[114]

In modern times, Anglo-Australian appellate courts have consistently left open the question.[115] This section seeks to answer it. It contains three parts. The first part considers the precedents and principled arguments supporting a 'cap'. The second part examines the cases and principled contentions favouring a 'no cap' view. The final part presents a principled approach to the issue that provides the only way of ensuring Anglo-Australian law makes sense, and is conceptually coherent.

I Cap

Carter and others argue that the 'balance of authority' favours the no cap view.[116] That is wrong. The majority of cases in fact support the view that a cap exists. Four example cases follow.

First, in *Wilbeam v Ashton*, Lord Ellenborough famously held: 'beyond the penalty you shall not go; within it, you are to give the party any compensation which [she] can prove [herself] entitled to'.[117] Second, in *Lord Elphinstone v Monkland Iron and Coal Co Ltd*, Lord Fitzgerald held that, where a penalty clause exists, A is entitled to recover 'the amount of [her] loss (not, however, exceeding the penalty) …'.[118] Third, in *Comr of Public Works v Hills*, delivering the Privy Council's advice Lord Dunedin held that A must be given an 'opportunity to prove such damages, not exceeding the sums in the penalties, as [she] can make out'.[119] Finally, in *Pearl Assurance Co Ltd v Government of the Union of South Africa*, the Privy Council held that, where a clause is held penal, A can recover her incurred loss 'but not exceeding the amount of the "penalty"'.[120]

Importantly, these statements are not conclusive. Some may be undermined. For instance, four years after *Wilbeam*, Lord Ellenborough CJ decided *Harrison v Wright*, in which

---

114 cf Andrew Burrows, *Remedies for Torts and Breach of Contract* (3rd edn, OUP 2004) 448 (cap) and Harvey McGregor, *McGregor on Damages* (19th edn, Sweet and Maxwell 2014) [15-027] (no cap).

115 *Cellulose Acetate Silk Co Ltd v Widnes Foundry (1925) Ltd* [1933] AC 20 (HL) 26; *Robophone Facilities Ltd v Blank* [1966] 1 WLR 1428 (CA) 1446; *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 174, 192, 212.

116 JW Carter and others, 'Contractual Penalties: Resurrecting the Equitable Jurisdiction' (2013) 30 JCL 99, 116.

117 (1807) 1 Camp 78, 78; 170 ER 883, 883.

118 (1886) 11 App Cas 332 (HL) 346.

119 [1906] AC 368 (PC) 376.

120 [1934] AC 570 (PC) 584.

his Lordship supported the no cap view.[121] Further, in *Watts, Watts and Co Ltd v Mitsui and Co Ltd*,[122] Lord Dunedin was one of the three Law Lords who approved Bailhache J's judgment in *Wall v Rederiaktiebolaget Luggude*.[123] As we shall see in a moment, the latter case supported the no cap view.

As to the principled arguments, the cap view has much support. Two points are considered below. The arguments, however, are refutable.

On the one hand, it is said that allowing uncapped recovery unjustifiably favours wrongdoers. For example, Hudson argues that insisting on a cap prevents placing a party who 'flouts the law' and demands a penalty in a better position than a party who 'observing the law' accepts a liquidated compensatory damages clause, as the latter is restricted to the sum fixed even if her incurred loss exceeds it.[124] However, as Chapter 4 illustrated, the Penalty Rule 'favours' wrongdoers by placing a breach requirement on the review jurisdiction, such that *only* a wrongdoer can engage the rule. This inconsistency eviscerates the writer's argument of any real force.

On the other hand, Burrows has submitted that, if no cap exists, A is actually *encouraged* to insert a penalty clause in her contract, as she can take advantage of the clause 'without suffering any disadvantage'.[125] In fact, the Penalty Rule encourages the use of penalty clauses because even if a clause is invalid, A is still permitted to recover from B her legally-recoverable loss. In short, even in the paradigm scenario,[126] there is no downside in A stipulating for a penalty clause. Accordingly, the writer's argument does not support a cap. Rather, it supports changing the *general* consequences of a court invalidating an agreed damages clause: on this view, in every case the Penalty Rule should ensure that A suffers some 'disadvantage' if the court invalidates the clause, thus discouraging the use of penalty clauses.

II No cap

A number of cases support the no cap view. Three examples follow.

---

121 (1811) 13 East 343, 347-48; 104 ER 402, 404.

122 [1917] AC 227 (HL) 245.

123 [1915] 3 KB 66 (KB).

124 AH Hudson, 'Penalties Limiting Damages' (1985) 101 LQR 480, 480.

125 Andrew Burrows, *Remedies for Torts and Breach of Contract* (3rd edn, OUP 2004) 448.

126 Text following n 20 in ch 1.

To start, in *Lowe v Peers*, Lord Mansfield held that A could elect to sue B in debt on the penalty clause, or proceed against B for damages for breach of covenant and 'recover more or less than the penalty *toties quoties*'.[127]

Next, several charterparty cases suggest that A may recover from B more than the penalty amount in general damages. For example, in *Wall*, a charterparty contained the following clause: 'Penalty for non-performance of this agreement proved damages, not exceeding estimated amount of freight.'[128] The shipowners refused to allow the charterers possession of the ship. Had the contract been performed, the estimated amount of freight that the charterers would have paid was £1125. The charterers incurred consequential losses of £3000. The shipowners argued that the clause was a limitation clause, limiting their liability to £1125. Conversely, the charterers contended that, as the clause was penal they could disregard it, and recover general damages for their incurred loss of £3000. Bailhache J held that his task was 'to decide between these two contentions'.[129] His Honour held that the charterers were right. In *Watts*, three members of the House of Lords—including Lord Dunedin—approved *Wall*.[130]

Lastly, in the Australian case of *W and J Investments Ltd v Bunting*, Lee J concluded that A is entitled to recover from B 'an amount equal to or greater than the amount referred to in the [penalty] clause',[131] citing Diplock LJ in *Robophone* in support.

Two of these example cases are unpersuasive.

One must begin with *Wall*, which is touted as the strongest authority in support of this view.[132] Bailhache J's reasons display four problems.

First, the judge took a false step, assuming that if the clause was not a limitation clause, it must have been a penalty clause. However, the parties' two contentions did not cover the field. The correct analysis is this. The clause may have been a limitation clause. Conversely, it may have been an agreed damages clause. In the latter case, if it was invalid, it was a penalty; if it was valid, it was a liquidated damages clause.[133] Accordingly, Bailhache J committed an

[127] (1768) 4 Burr 2225, 2228; 98 ER 160, 163.

[128] *Wall v Rederiaktiebolaget Luggude* [1915] 3 KB 66 (KB) 71.

[129] *Wall v Rederiaktiebolaget Luggude* [1915] 3 KB 66 (KB) 71.

[130] *Watts, Watts and Co Ltd v Mitsui and Co Ltd* [1917] AC 227 (HL) 235, 245, 246.

[131] (1984) 1 NSWLR 331 (NSWSC) 336.

[132] Harvey McGregor, *McGregor on Damages* (19th edn, Sweet and Maxwell 2014) [15-093].

[133] A true dichotomy exists between a liquidated damages clause and a penalty clause: there is 'no half-way house': *Bridge v Campbell Discount Co Ltd* [1962] AC 600 (HL) 633; *Coates v Sarich* [1964] WAR 2 (WASCFC) 13.

'error of logic'[134]: the judge ignored the possibility that the clause might have been a liquidated compensatory damages clause.

Second, if properly classified the clause was an agreed damages clause, arguably it was not penal. Bailhache J's reasoning regarding why the clause was penal does not withstand scrutiny. To begin with, the judge relied heavily on the history of the particular clause in charterparties. Its traditional form was as follows: 'Penalty for non-performance of this agreement estimated amount of freight'. The original form of this clause was said to be 'obviously' a penalty clause,[135] on the basis that it failed the Indiscriminate Cause Test.[136] Bailhache J held that the addition of four words—'proved damages, not exceeding'—had not changed the clause's nature. In particular, the judge said that commercial parties familiar with the usual charterparty clauses 'do not read apparently common form clauses carefully'.[137] Accordingly, a shipowner needed to use much clearer language to bring home to the charterers the change in effect to the original form of clause. This argument has purchase is the *consumer* context, but it cannot be correct in the *commercial* context. If a commercial party fails to read a document carefully, and obtain proper advice regarding its effect, then it must bear the consequences. Carelessness is no excuse. Further, Bailhache J placed excessive weight on the parties having described the clause as a 'penalty'. Although relevant, as it may reverse the burden of proof,[138] this fact should not be so influential.

Third, if properly classified the clause *was* an agreed damages clause, then arguably it was a liquidated damages clause. The court may have construed the words 'non-performance of *this agreement*' as limiting the clause to cases of complete non-performance. If so, the clause only applied where either party substantially failed to perform the contract. Consequently, the clause would not have failed the Indiscriminate Cause Test.

Finally, if properly classified the clause was *not* an agreed damages clause, then it was in fact a limitation clause. The clause is best interpreted in this way. The form of words used clearly provided for a limitation clause, and not an agreed damages clause. The use of the words 'not exceeding' suggests a *limit* on, not a *setting* of, the amount recoverable. Also, the use of the phrase 'proved damages' meant the clause required proof of incurred loss up to and including the estimated amount of freight, such that if the charterers' loss were *less* than the

---

[134] Michael Furmston (gen ed), *The Law of Contract* (4th edn, Butterworths 2010) [8.126] fn 3.

[135] *Wall v Rederiaktiebolaget Luggude* [1915] 3 KB 66 (KB) 71.

[136] *Widnes Foundry (1925) Ltd v Cellulose Acetate Silk Co Ltd* [1931] 2 KB 393 (CA) 408, affd [1933] AC 20.

[137] *Wall v Rederiaktiebolaget Luggude* [1915] 3 KB 66 (KB) 74.

[138] *Willson v Love* [1896] 1 QB 626 (CA) 630.

estimated amount of freight, they would only recover that amount. In *Watts*, Lord Finlay LC recognised this, stating that if the clause 'had appeared for the first time … it might have been construed as imposing a limitation on the damages to be recovered'.[139]

In short, it has been held that the charterparty cases are exceptional, turning 'largely upon the history of the particular clause'.[140] They also turn upon questionable reasoning. They are unreliable.

Turning to *Bunting*, two points are noteworthy. For one, the comment was obiter dicta, since the penalty amount far exceeded the incurred loss.[141] For another, Lee J read *Robophone* incorrectly. The judge wrongly viewed the case as supporting the proposition that general damages permissibly may exceed the penalty amount. In truth, Diplock LJ's dictum was equivocal.[142] This case, then, is also unreliable.

We must move to the principled arguments for aid. Two such arguments appear to support the no cap view. As will be demonstrated, the first argument is problematic. Conversely, the second argument is logically unimpeachable.

The first argument is that the 'mutuality principle' demands that no cap exist. Greig and Davis propound this argument in the following way. If the clause is penal and A's incurred loss is *less* than the penal amount, the clause cannot assist A, and A must prove her loss. It is said to follow that, if the clause is penal and A's incurred loss is *more* than the penal amount, the clause cannot assist B. Greig and Davis contend that the 'necessary mutuality' between the parties would be destroyed if 'the loss being greater than the penal sum, [B] were able to rely on the clause to limit his liability'.[143] The basis of this argument is misconceived. To explain, assume a cap exists. If so, whether an agreed compensatory damages clause operates to A's advantage is properly determined by whether it is valid, or not. On the one hand, where the clause is *valid*, it operates to A's advantage. For example, if the clause is valid and A's incurred loss is less than the liquidated compensatory damages amount, the clause advantages A, because B cannot restrict A to her lower incurred loss. On the other hand, where it is *invalid*, it operates to B's advantage, as in the premise B can restrict A's recovery to

---

[139] *Watts, Watts and Co Ltd v Mitsui and Co Ltd* [1917] AC 227 (HL) 235.

[140] *Widnes Foundry (1925) Ltd v Cellulose Acetate Silk Co Ltd* [1931] 2 KB 393 (CA) 407, affd [1933] AC 20 (HL).

[141] *W and J Investments Ltd v Bunting* (1984) 1 NSWLR 331 (NSWSC) 334, 343.

[142] See n 115.

[143] DW Greig and JLR Davis, *The Law of Contract* (Lawbook Company 1987) 1452.

the penal sum. When evaluated along this axis—of contractual validity—perfect 'mutuality' exists.

Further, it is more theoretically sound to adopt an evaluative axis of contractual validity, rather than the amount of A's loss in cases of invalidity. The first basis for distinction has normative force: the fact the agreed damages clause is valid in one case, and invalid in the other, provides a normative justification for treating A and B differently in both scenarios. The second basis for distinction has little, if any, normative power: the fact that A has suffered more loss in one case, and less in another, is normally a matter of chance, and it does not provide a justificatory basis for treating the parties differently in both scenarios.

The second argument is this. Beatson, Burrows and Cartwright argue that where a clause is held penal, 'the damages recoverable must be assessed in the usual way', and that 'it is possible for the claimant to recover a sum … that is greater than the stipulated penal sum'.[144] This reasoning implicitly appears to apply the 'expunction' approach. The reasoning process seems to be as follows: if a penalty clause is invalidated, it is properly void, and stuck out of the contract; as the agreed damages regime is a nullity, the court's jurisdiction to determine A's general damages is reinstated; as the court treats the clause as non-existent and disregards it for all purposes, it cannot logically operate to restrict a general damages award. It follows that A must be able to recover her agreed compensatory damages without limit. No cap exists.

Given this, Peel has argued that a cap view is flawed because under it an invalid penalty clause impermissibly metamorphoses into a valid limitation clause. On this view, a clause that was 'originally invalid' as a penalty clause will have 'become valid as a limitation clause simply by reason of a change of circumstances'.[145] The reasoning also appears to apply the 'expunction' approach, but adds two additional steps. At the time the general damages regime is reinstated, A's incurred loss is more than the penal amount; B then 'revives' and relies upon the formerly invalid clause to limit his general damages liability to the penal amount, such that the clause has become a valid limitation clause *merely* due to an *ex post* consideration. This argument is unassailable.

III A principled approach

On this issue, Anglo-Australian private law appears incoherent. The cases seemingly support the view that a cap exists, such that the penalty amount caps A's right to general compensatory

---

[144] Jack Beatson, Andrew Burrows and John Cartwright, *Anson's Law of Contract* (29th edn, OUP 2010) 571.

[145] Edwin Peel, *The Law of Contract* (13th edn, Sweet and Maxwell 2011) [20-142].

damages. However, the principled arguments show that this view creates inconsistency and leads to illogicality. As to the former, the cap view suggests that penalty clauses are in fact void. This, however, contradicts Anglo-Australian law as it presently stands. As to the latter, once a court declares a penalty clause invalid, it apparently exists in a twilight zone: it is for all purposes dead, but it then can be resuscitated for a new purpose, depending on how events play out after the contract date. This is an extremely unattractive position that we should resist accepting.

What is required is a principled approach to the cap issue. Anglo-Australian private law needs an approach that helps us to render the law on this issue intelligible and coherent. Such an approach follows. It comprises five points.

First, if a penalty clause is held invalid, Anglo-Australian private law takes the 'compression' approach.

Second, pursuant to the compression approach, a penalty clause is partially unenforceable. In other words, the clause is unenforceable *pro tanto*, that is, to the extent that the penal sum exceeds A's incurred loss. Per hypothesis, A's incurred loss is more than the penalty amount. Accordingly, A may recover for her incurred loss by enforcing the clause to its full extent, but no more. B does not rely on the clause to limit his liability; the clause merely does so by dent of the fact that it remains in the contract. The clause does not undertake any change in character, and thus does not become a valid limitation clause.

Third, under the compression approach, the penalty clause is 'compressed' to the extent to which it is enforceable. However, it remains in the contract, and 'covers the field' regarding the compensatory response to a breach falling within its scope. Thus, the agreed damages regime subsists, standing in the place of the general damages regime. The clause prevents A from bringing a general compensatory damages claim at all. So, it is incorrect to say that, as the penalty clause remains in the contract, A is entitled 'to claim general damages subject to a cap' at the value of the penal sum.[146] Rather, a penalty clause cannot operate as a cap on A's general compensatory damages claim because, in fact, she can bring so such claim in the first place. The cap issue, then, is a red herring.

Fourth, given that A has no right to claim general compensatory damages, the cases that seemingly take the cap view are best interpreted as ones where A enforced the penalty clause to the extent of her legally-recoverable loss. Read correctly, the cases illustrate that A

---

[146] Finola O'Farrell, 'Challenging and Defending Liquidated Damages' (2004) 56 Amicus Curiae 20, 23.

may partially enforce a penalty clause up to, at a maximum, the penalty amount. Put another way, the perceived award of general compensatory damages subject to a cap is the enforcement of the penalty clause in disguise.

Finally, given that the case law all properly deals with A's enforcing the penal clause, it follows that, as best understood, Anglo-Australian courts *only* partially enforce the relevant clause. They do not award general compensatory damages where a clause is penal. Therefore, A has no right to elect between two means of recovery.

One last point is noteworthy. On the principled approach, the only possible consequence of a court invalidating a penalty clause is that A's right of action is truncated, such that A is only entitled to enforce the clause up to and including the amount of her incurred loss. If so, then *practically* it does not matter whether a derisory agreed compensatory damages clause is valid or not. If the clause is valid, A's right of action is truncated,[147] and the derisory liquidated compensatory damages clause will *set* the exact amount that A may recover. A will only recover the (inadequate) liquidated compensatory damages sum. If the clause is invalid, A's right of action is again truncated, although the penalty clause will *limit* the amount that A may recover. However, A will only recover the same sum: although in the premise A has suffered more loss, her recovery is restricted to the penalty sum, as logically one can only enforce a penalty clause up to its uppermost figure. Accordingly, in either case, the clause will have the same commercial effect.

In conclusion, the principled approach is attractive because it provides an explanation that makes sense of the case law, and is intellectually coherent. Thus, the principled approach best explains the law as it stands.

3 Conclusion

This chapter began by examining the consequences of a court validating a clause as a liquidated damages clause.

The chapter first considered the true nature of liquidated damages. It illustrated that such a concept generally is in the nature of a debt, not damages. The chapter then analysed how liquidated *compensatory* damages clauses operate in two situations. Where A's incurred loss is *less* than the liquidated compensatory damages amount, she may enforce the clause nonetheless. Conversely, where A's incurred loss is *more* than the liquidated compensatory

[147] Text following n 119 in ch 7.

damages amount, a liquidated compensatory damages clause provides an exhaustive remedy for compensatory damages for B's default. However, A may recover from B additional, general compensatory damages for an independent kind of loss arising from a breach of a distinct duty.

The chapter then considered three issues concerning penalty clauses.

To start, as to the true nature of a penalty, the chapter established that properly viewed penalties generally are akin to a debt; and that a penalty clause is partially unenforceable, not void.

Next, the chapter analysed the effect of a penalty clause when A's incurred loss is *less* than the penalty amount. It reached two important conclusions. For one, although it appears that A may elect either to institute against B a general damages claim or sue B on the penalty clause, in truth, no right to elect exists. The Penalty Rule is best understood as providing A with a right of action on the penalty clause. This view is transparent, and renders the rule theoretically sound. For another, although this situation raises three difficult theoretical issues, they are resolvable.

Lastly, the chapter considered how a penalty clause operates when A's incurred loss *exceeds* the penalty sum. The most pressing question in this context was this: if A may sue B for breach of contract, can A claim against B general compensatory damages in the ordinary way, or does the penal sum 'cap' her recovery? The chapter exposed this question as a red herring. The penalty amount cannot limit A's right to recover general compensatory damages, for no such right exists. Rather, if a clause is penal, it is compressed to the extent of its enforceability, and A has the power to sue on the clause to recover. When A's incurred loss exceeds the penal sum, no compression is required. A may sue on the clause to recover the entire penal amount.

At this point, it is important to note that the first aim of this thesis—set out in Chapter 1—has been achieved. We have identified and understood the Penalty Rule's proper scope and content; we have restated the law. The question then arises as to whether the restated law is justifiable. The next part turns to consider that question.

Part Two: Justification

7 THEORIES

There are as many putative theoretical justifications of the Penalty Rule as there are scholars writing on the topic. As such, any examination of all the theories on offer would be unwieldy, and unacceptably long. Accordingly, this chapter evaluates the *leading* theories.

This chapter considers the Penalty Rule's purported theoretical justifications. Three prominent theories are discussed. First, the chapter considers Chen-Wishart's substantive unconscionability explanation for the rule. Second, the chapter analyses the argument of many law-and-economics scholars that the doctrine in truth promotes efficiency. Finally, it places the orthodox view—that the rule's true rationale is to prevent 'overcompensation'—under the microscope. In line with the interpretive methodology of this thesis, this chapter adopts the criteria of fit, transparency and coherence to evaluate the three potential theoretical justifications of the rule. In addition, the chapter pays close attention to the principled arguments supporting or undermining a particular theory, and any relevant practical implications of a specific rationale.

The chapter illustrates that each proposed theory is, to a greater or lesser extent, flawed. This explains why many judicial comments on the rule's justification involve a counsel of despair. It also explains why eminent scholars have called for the Penalty Rule to be expunged from the private law. Prima facie, this demand seems justified. Unless a robust, coherent understanding of the rule is found, it should be cast aside.

1 Substantive unconscionability

A Thesis

Two forms of unconscionability arguably exist. The first is *procedural* unconscionability, which is concerned with fairness in the transactional process. The second is *substantive* unconscionability, which focuses on fairness in the transactional outcome. Many writers have argued that procedural unconscionability theoretically justifies the Penalty Rule.[1] The leading unconscionability theory, however, is the substantive unconscionability theory insightfully articulated by Chen-Wishart.[2] It is examined in this section.

Chen-Wishart argues that substantive unconscionability is the proper theoretical basis of the Penalty Rule: the concept both explains and informs the rule. This rationale is premised upon the fact that the test of validity focuses on the imbalanced *result* of the parties' agreement concerning agreed damages.

The writer argues that 'substantive unconscionability' means 'unfairness' or 'undue harshness'. As to the former, Chen-Wishart states that the rule's proper basis lies in preventing the 'unfairness in … the resulting exchange' between the parties.[3] As to the latter, it is said that penalty clauses are not enforced because the lop-sided outcome may 'entail undue harshness to contract-breakers',[4] regardless of how that outcome was procured, and even if the parties willingly agreed to the clause.

After contending that substantive unconscionability 'is and must be the underlying basis' of the Penalty Rule, Chen-Wishart focuses her attention on the results of recognising the doctrine's alleged proper rationale.[5] The writer argues that the rule must be 'realigned' with its substantive unconscionability rationale, and identifies several ways of achieving this. She first proposes a realigned jurisdictional test, and then proffers a realigned test of validity.

As to jurisdiction, the writer argues that the breach requirement is unsatisfactory, as it is rigid and artificial. It should be abandoned. Instead, the court's review power should be 'confined to specifically delimited high risk areas' of substantive unfairness.[6] Chen-Wishart notes that problem of crafting a jurisdictional test based on substantive unconscionability that

---

[1] eg Tony Downes, 'Rethinking Penalty Clauses' in Peter Birks (ed), *Wrongs and Remedies in the Twenty-First Century* (Clarendon Press 1996) 260, 263-65, 267-68; Elisabeth Lanyon, 'Equity and the Doctrine of Penalties' (1996) 9 JCL 234, 240-42.

[2] See also DJ Ibbetson, *A Historical Introduction to the Law of Obligations* (OUP 2001) 255-56; GE Dal Pont, *Equity and Trusts in Australia* (5th edn, Lawbook Company 2011) [13.10].

[3] Mindy Chen-Wishart, 'Controlling the Power to Agree Damages' in Peter Birks (ed), *Wrongs and Remedies in the Twenty-First Century* (Clarendon Press 1996) 283.

[4] Mindy Chen-Wishart, *Contract Law* (4th edn, OUP 2012) [14.3.1.1].

[5] Mindy Chen-Wishart, 'Controlling the Power to Agree Damages' in Peter Birks (ed), *Wrongs and Remedies in the Twenty-First Century* (Clarendon Press 1996) 283.

[6] Mindy Chen-Wishart, 'Controlling the Power to Agree Damages' in Peter Birks (ed), *Wrongs and Remedies in the Twenty-First Century* (Clarendon Press 1996) 298.

does not collapse into a general unfairness review. In response, the writer advocates 'Sustainable evolution and not revolution …'.[7]

As to validity, Chen-Wishart explicitly refrains from providing a precise test of substantive unfairness. The courts 'must be left to develop the specifics on a case-by-case basis'.[8] However, the writer argues that a doctrine that embraces its true basis will take account of particular factors. Four are noteworthy. First, courts should only invalidate an agreed damages clause where the unconscionability is 'manifest'. Second, the test of validity should consider the contract holistically. For example, courts should take account of any 'premium' that A has paid to B in exchange for an 'inflated' agreed damages clause. Third, it is imperative that the test of validity takes account of A's incurred loss. Finally, judges should consider the term's subjective impact on the parties, including B's financial circumstances and his ability to pay the agreed damages.

Chen-Wishart concludes that her realigned model will result in less judicial intervention and a 'more finely calibrated justice'.[9]

Finally, the writer argues that the substantive unconscionability rationale is both transparent and coherent.

On the one hand, Chen-Wishart asserts that the theory aligns with the courts' general reasoning in penalty clause cases. The writer provides examples. One is *Protector Endowment Loan and Annuity Co v Grice*, in which Bramwell LJ held that in Penalty Rule cases 'equity in truth refused to allow to be enforced what was considered to be an unconscientious bargain'.[10] Another is *AMEV-UDC Finance Ltd v Austin*, where Mason and Wilson JJ held that the Penalty Rule operated to 'relieve against provisions so unconscionable or oppressive that their nature is penal rather than compensatory'.[11]

On the other hand, Chen-Wishart argues that openly recognising the Penalty Rule's unconscionability rationale would make the private law strongly coherent. The writer states that unconscionability 'can provide a unifying concept for a number of distinct rules dealt with

---

[7] Mindy Chen-Wishart, 'Controlling the Power to Agree Damages' in Peter Birks (ed), *Wrongs and Remedies in the Twenty-First Century* (Clarendon Press 1996) 298.

[8] Mindy Chen-Wishart, *Contract Law* (4th edn, OUP 2012) [9.6.2].

[9] Mindy Chen-Wishart, 'Controlling the Power to Agree Damages' in Peter Birks (ed), *Wrongs and Remedies in the Twenty-First Century* (Clarendon Press 1996) 298.

[10] (1880) 5 QBD 592 (CA) 596.

[11] (1986) 162 CLR 170 (HCA) 193.

under different headings …'.[12] Accordingly, it is said that the rule should be viewed as one doctrine in a unified series along with certain vitiating factors, such as duress and undue influence, and powers of relief against certain 'remedial' terms, such as deposit clauses and part payment forfeiture clauses. Thus, the Penalty Rule is one instance of a more general phenomenon whereby the private law interferes with the parties' bargain based on a procedural or substantive 'defect'. Accordingly, this theory seemingly scores well on the coherence criterion for assessing private law theories.

B Evaluation

Upon close inspection, Anglo-Australian private law does not countenance a substantive unconscionability rationale for the Penalty Rule. This proposed theoretical justification raises issues of fit, transparency, coherence, principle and practice. They are considered in order.

I Fit

For three reasons, the substantive unconscionability theory scores poorly on the fit criterion.

First, Chen-Wishart's contends that substantive unconscionability 'is and must be' the Penalty Rule's underlying basis. Her attempt to justify this contention, however, is flawed. For example, the writer states that applying the tests of validity as at the contract date 'makes sense if the concern is fairness but not if it is strict fidelity to the compensatory measure'.[13] However, it is said that the substantive unconscionability rationale requires that the rule take account of A's incurred loss, which is necessarily evaluated at the judgment date. It follows that, on the rationale's own terms, it cannot explain an *ex ante* assessment date. Further, Chen-Wishart states that a fairness rationale is consistent with enforcing an agreed damages clause that passes the Impossibility Test.[14] However, this is a mere assertion. To assess whether the substantive unconscionability theory fits this particular test of validity, we need to be told what 'fair' means and then shown *why* a clause that passes this test is fair. We are not told the former, nor shown the latter.

Second, the most important feature of this theory is what is absent from it. That is, the substantive unconscionability rationale does not attempt to justify critical features of the

---

[12] Mindy Chen-Wishart, *Contract Law* (4th edn, OUP 2012) [9.6.2].

[13] Mindy Chen-Wishart, 'Controlling the Power to Agree Damages' in Peter Birks (ed), *Wrongs and Remedies in the Twenty-First Century* (Clarendon Press 1996) 278.

[14] Mindy Chen-Wishart, 'Controlling the Power to Agree Damages' in Peter Birks (ed), *Wrongs and Remedies in the Twenty-First Century* (Clarendon Press 1996) 278.

Penalty Rule. Accordingly, the writer does not demonstrate that the theory is consistent with, and supported by, these aspects of the rule. So, the purported rationale does not articulate why the breach requirement exists, or why the paradigm scenario[15] involves a breach. As Furmston points out, 'this is an odd rule to have if the true basis of the penalty distinction is unfairness …'.[16] Also, the proposed justification does not explain why a derisory agreed compensatory damages clause is valid. In fact, judges[17] and staunch unconscionability supporters[18] have conceded that such a clause is unconscionable, and should be penal. In addition, the substantive unconscionability thesis simply does not tell us why penalty clauses are partially enforced after the court has suppressed the penal element. Indeed, the usual remedy for 'unconscionability' is rescission of the *entire contract*. In short, the substantive unconscionability rationale is inadequate.

Finally, Chen-Wishart's proposed 'realigned' model is unsound.

Consider first the realigned review jurisdiction. A substantive unconscionability approach eschews the breach requirement, but it then cannot delimit the court's review power. There are two points here.

On the one hand, although the writer refers to limiting the review jurisdiction to particular 'high risk areas', she fails to delineate any such areas in the commercial sphere. In fact, the writer specifically rejects two potential situations, those displaying a risk of cognitive error and the absence of bargaining.[19]

On the other hand, any substantive unconscionability concept appears unable *logically* to delimit the sphere of review in a theoretically satisfactory manner. As Chen-Wishart states, 'in principle, concern about substantive unfairness attaches to all terms of a contract and not just to agreed damages [clauses]'.[20] Hence, unconscionability is inadequate in terms of pure theory, because it is not able to achieve a central goal of a theory of the relevant kind: to reveal the review jurisdiction's appropriate scope. In short, this rationale is unable rationally to confine the jurisdictional test to prevent the review of *all* clauses involving an obligation to pay.

---

15 Text following n 20 in ch 1.

16 MP Furmston, 'Contract Planning: Liquidated Damages, Deposits and the Foreseeability Rule' (1991) 4 JCL 1, 8.

17 *Silent Vector Pty Ltd v Squarcini* [2008] WASC 246, (2009) 25 BCL 58 [63].

18 eg SM Waddams, *The Law of Contracts* (6th edn, Canada Law Book 2010) [459].

19 Mindy Chen-Wishart, 'Controlling the Power to Agree Damages' in Peter Birks (ed), *Wrongs and Remedies in the Twenty-First Century* (Clarendon Press 1996) 290-92.

20 Mindy Chen-Wishart, 'Controlling the Power to Agree Damages' in Peter Birks (ed), *Wrongs and Remedies in the Twenty-First Century* (Clarendon Press 1996) 298.

The writer implicitly concedes the point, acknowledging the problem but failing to offer any solution, preferring elliptically to advocate 'evolution and not revolution'.

Now consider three aspects of the realigned test of validity.

To start, the substantive unconscionable rationale requires that courts take a holistic approach. However, if an agreed compensatory damages clause fails the test of validity, it will be held penal 'even if the contract price had been doubled or tripled to compensate for the provision'.[21] In essence, A's payment to B of a 'premium' is properly irrelevant. Courts do not take a 'holistic' approach; they concentrate solely on the alleged penalty clause.[22]

Next, Chen-Wishart's model advocates courts taking account of *ex post* events, such as A's incurred loss. This is troubling. It flatly contradicts established authority.[23] Further, as each party will not know in advance what A's right, and B's duty, will be upon B's default, it will eliminate one of the main advantages of liquidated damages clauses—namely, certainty.[24]

Lastly, the writer suggests that courts should consider the clause's subjective impact on the parties, for example B's financial circumstances. However, the English Court of Appeal has insisted that the parties' financial circumstances are immaterial.[25] In any event, in the private law generally, impecuniosity is not a defence.

II Transparency

Procedural and substantive unconscionability are heavily interrelated, so much so that Atiyah questions whether any distinction between the two actually exists.[26] Chen-Wishart seemingly concedes this, concluding that where substantive unfairness is manifest, courts should grant relief either because procedural unconscionability can be inferred *or* on the 'more direct basis' of substantive unconscionability.[27] Accordingly, it is ambiguous whether the writer is arguing that a defect in the transactional process or outcome is the true ground for intervention. This creates a transparency issue. Given it is unclear whether the contention is that procedural or

---

[21] Stephen A Smith, *Atiyah's Introduction to the Law of Contract* (6th edn, Clarendon Press 2005) 308.

[22] *Jeancharm Ltd v Barnet Football Club Ltd* [2003] EWCA Civ 58 [29].

[23] eg *Comr of Public Works v Hills* [1906] AC 368 (PC) 376; *Hamilton v Lethbridge* (1912) 14 CLR 236 (HCA) 264.

[24] Text accompanying and following n 16 in ch 1.

[25] *Imperial Tobacco Co (of Great Britain and Ireland) Ltd v Parslay* [1936] 2 All ER 515 (CA) 523, 526. See also *Citicorp Australia Ltd v Hendry* (1985) 4 NSWLR 1 (NSWCA) 23.

[26] PS Atiyah, *Essays on Contract* (Clarendon Press 1990) 333-35.

[27] Mindy Chen-Wishart, 'Controlling the Power to Agree Damages' in Peter Birks (ed), *Wrongs and Remedies in the Twenty-First Century* (Clarendon Press 1996) 297.

substantive unconscionability is the true basis for the court's intervention, we cannot be confident as to what is the precise theory. Consequently, we simply cannot tell whether the theory lines up with the court's reasoning in Penalty Rule cases.

If, properly understood the writer supports a substantive unconscionability rationale, another issue of transparency arises. Anglo-Australian courts have rejected substantive unconscionability as the specific basis for the Penalty Rule. Courts have consistently held that they have no jurisdiction to rewrite a term of parties' contract because they view it to be 'unduly onerous'.[28] It is said that the rule is not designed to relieve a party from an 'onerous' or 'commercially imprudent' bargain.[29] In Lord Radcliffe's words, 'unconscionable' cannot be 'a panacea for adjusting any contract between competent persons when it shows a rough edge to one side or the other …'.[30] Accordingly, the substantive unconscionability theory does not sufficiently take account of the reasons that judges give for their decisions in this context.

III Coherence

Two issues arise as regards the coherence criterion, as follows.

For one, the substantive unconscionability theory does not satisfy the weak notion of coherence, for it would reveal contract law to be internally inconsistent. Contract law is not interested in any inequality of values exchanged. For example, an inadequacy of consideration does not invalidate a contract.[31] If substantive unconscionability properly justified the Penalty Rule, then the law here would be said to prevent 'unfairness in … the resulting exchange'.[32] Accordingly, in the premise, contract law would contradictorily dismiss 'manifest unfairness in the contract price', and strictly police 'manifest unfairness in the remedy'.[33]

For another, the substantive unconscionability rationale does not satisfy the strong coherence notion, for properly understood it cannot unify the private law, or parts of it, under a single theme. Thus, Chen-Wishart argues that a principle of restraining substantive unconscionability theoretically justifies several contract law rules, including the Penalty Rule,

---

[28] *Robophone Facilities Ltd v Blank* [1966] 1 WLR 1428 (CA) 1446.

[29] *Export Credits Guarantee Department v Universal Oil Products Co* [1983] 1 WLR 399 (HL) 403.

[30] *Bridge v Campbell Discount Co Ltd* [1962] AC 600 (HL) 626.

[31] *Chappell and Co Ltd v Nestlé Co Ltd* [1960] AC 87 (HL) 108.

[32] Mindy Chen-Wishart, 'Controlling the Power to Agree Damages' in Peter Birks (ed), *Wrongs and Remedies in the Twenty-First Century* (Clarendon Press 1996) 283.

[33] Sarah Worthington, *Equity* (2nd edn, OUP 2006) 220.

and that this principle instantiates a distributive justice norm.[34] In this thesis, 'distributive justice' refers to the Aristotelian form of justice concerned with the overall distribution of a divisible good, such as wealth, among the members of a community. Distributive justice divides the good in accordance with some criterion that compares the member's relative merits. There are many criteria for assessing the justice of any distribution: for example, need or skill. So, distributive justice embodies a 'proportional equality'.[35] Accordingly, Chen-Wishart seemingly argues that preventing substantive unconscionability in contract law helps to create or maintain a just distribution of society's resources. In particular, judges have the power to invalidate penalty clauses to redistribute wealth. Thus, on this view, the Penalty Rule impeaches some agreed damages clauses because 'they offend [an] important societal interest that overrides the usual priority given to personal autonomy …'.[36]

There are two responses here. Positively, the Penalty Rule does not instantiate distributive justice because it instantiates corrective justice. The next chapter will show this. Negatively, if the rule *did* instantiate distributive justice, this would in fact create *incoherence*, as Anglo-Australian contract law does not instantiate the distributive justice norm. This is a controversial philosophical issue. Therefore, this thesis cannot discuss it fully. Nevertheless, the following three comments support the proposition.

To start, Anglo-Australian courts generally are not empowered to compel strangers to enter contracts. This severely limits the *opportunity* for contract law to redistribute wealth. Even if contract law had sufficient opportunities to redistribute wealth, it would not have the *ability* to do so satisfactorily. Contract law does not conduct wealth assessments or wealth comparisons—either as between the parties, or between one party and the relevant members of society.

Next, in many cases contract law does not justly distribute society's resources; it actually does not affect the distribution at all. So, where a contract is set aside, the law will merely restore the *status quo ante*.[37] However, most concede that the pre-existing state of affairs is distributively unjust. For example, most parties are of unequal wealth before a contract is made. Consequently, in ordering rescission, the court merely reinstates an unjust distribution. This may not make a poor party worse off, but it will not make her better off.

---

34 Mindy Chen-Wishart, *Contract Law* (4th edn, OUP 2012) [1.4.4], [11.10], [14.3.2.2].

35 Ernest J Weinrib, 'Corrective Justice in a Nutshell' (2002) 52 UTLJ 349, 349.

36 Sarah Worthington, *Equity* (2nd edn, OUP 2006) 219.

37 An alternative analysis is that it is the law of unjust enrichment, and not contract law, that reinstates the pre-existing state of affairs.

Lastly, in other cases where contract law does affect societal distributions, it in fact aggravates distributional inequalities. In most legally significant, commercial contracts, the parties exchange rights of unequal value. So, where a contract is enforced or general compensatory damages are ordered, contract law will bring about an unequal exchange. In many cases, the contract terms will favour the wealthy party. If so, the poor party will be left worse off. Clearly, then, contract law does not regulate individuals' behaviour for the common good.

In short, the unconscionability thesis is flawed because it fails to render contract law either weakly, or strongly, coherent.

IV Principle

For Chen-Wishart, 'substantive unconscionability' means either 'undue harshness' or 'unfair'. Further, the writer propounds that courts must work out its constituent principles on a 'case-by-case' basis. Each argument is problematic.

First, an 'undue harshness' definition is innately defective. It suggests that a 'harsh' outcome is acceptable, provided that it is not 'undue'. However, the assertion that an outcome is 'undue' simply 'begs the question', in that it assumes the truth of the proposition advanced, without arguing for it. It does so because although substantive unconscionability theory purports to be based upon rendering to A only that which she is due, it does not explain that which is properly due. In other words, the theory does not explain *how* the private law should calculate the proper outcome. This leaves open the possibility that A is in fact entitled to the particular outcome. If so, then the outcome cannot be 'unduly' harsh.

Second, 'fairness' is a vague concept. As Beever notes, 'Fairness in the round is too abstract a principle, and its content too uncertain … sufficiently to direct decision making'; the fairness concept is objectionable because 'it is so wide open that it does not sufficiently direct judicial decision making and is so controversial that no consensus on its application could be expected'.[38] Accordingly, abstract 'unfairness' is unworkable as a standard for invalidating rights and obligations, and for legal adjudication. Judges agree with these views. For example, writing extra-curially, McHugh J notes that 'When courts have to apply vague standards, consistency of decision-making—which is one of the primary benefits of the rule of law—is

[38] Allan Beever, *Rediscovering the Law of Negligence* (Hart Publishing 2007) 50.

difficult to achieve. Moreover, the decision of a court applying such vague criteria often seems arbitrary.'[39]

Admittedly, Chen-Wishart does offer some particular 'factors' to consider in determining the substantive fairness of an agreed damages clause. However, to the question of what counts as 'substantive fairness', there is no necessarily correct criteria to apply to determine the issue, and no necessarily right answer. Evaluations of this kind, based on a nebulous concept with no demonstrably correct outcome one way or the other, do not credit the private law.

Finally, Chen-Wishart refrains from expounding her notion of substantive unconscionability, adding that much of it 'will have to be worked out from the experience of cases'.[40] Similarly, some judges have held that courts should work out the content of the 'unconscionability' concept on a case-by-case basis. For example, in *National Westminster Bank plc v Morgan*, Lord Scarman held that 'Definition is a poor instrument' to ascertain unconscionability, for it is 'a question which depends upon the particular facts of the case'.[41] These views suggest that decisions about what constitutes 'substantive unconscionability' will turn on the facts of the particular case. If this interpretation is correct, the argument must be rejected. Courts' role in expounding legal principle is to establish how and why the circumstances of the case result in liability, or not. It does not suffice for a judge to scrutinise the facts—however closely—and then conclude that the defendant is, or is not, legally responsible. Nor will it suffice for a judge to articulate 'principles' that are incapable of illuminating the decision-making process that led to the decision. This approach, in truth, entails substituting palm tree justice for the rule of law.

V Practice

The unconscionability theory raises two practical concerns.

Practitioners judge the practical value of a commercial law rule by whether it is achieves certainty and predictability. As to the former, Chen-Wishart confesses that any notion of 'unconscionability' will involve an 'inevitable element of uncertainty and discretion',

39 Michael McHugh, 'The Growth of Legislation and Litigation' (1995) 69 ALJ 37, 43.

40 Mindy Chen-Wishart, 'Controlling the Power to Agree Damages' in Peter Birks (ed), *Wrongs and Remedies in the Twenty-First Century* (Clarendon Press 1996) 296.

41 [1985] AC 686 (HL) 709. See similarly *Commonwealth of Australia v Verwayen* (1990) 170 CLR 394 (HCA) 440-41.

and that her proposed realignment will make the Penalty Rule more uncertain.[42] Indeed, judges have 'exposed, and deprecated' the uncertainty inherent in the unconscionability concept.[43] As to the latter, the fact 'unconscionability' lacks a sufficiently certain content means the outcomes of cases relying it are necessarily unpredictable. An unconscionability rationale, then, is practically of little value.

Also, as Lord Nicholls said in *Royal Brunei Airlines Sdn Bhd v Tan*, the word 'unconscionable' is 'not a word in everyday use by non-lawyers'.[44] Thus, Shelton notes the practitioner view that a substantive unconscionability notion is commercially incomprehensible: 'Tell a commercial client that [she] must not act unconscionably in agreeing [her] contract and [she] won't know what you are talking about …'.[45] An explanation in unconscionability terms will at best not aid commercial understanding, and at worst hinder it.

VI Conclusion

Clearly, substantive unconscionability is not the Penalty Rule's proper juristic basis. In this context as in many others, the term 'unconscionable' may have 'masked rather than illuminated the underlying principles at stake'.[46] The Penalty Rule, then, is better explained on other grounds.

2 Efficiency theory

A Thesis

Much economic literature examines the Penalty Rule. Arguably, the neo-classical school of law-and-economic analysis has been the most influential. Posner is its most distinguished proponent. Accordingly, this section will focus on the perceptive arguments that Posner has raised.

Three introductory points arise.

---

[42] Mindy Chen-Wishart, 'Controlling the Power to Agree Damages' in Peter Birks (ed), *Wrongs and Remedies in the Twenty-First Century* (Clarendon Press 1996) 284, 298.

[43] *Mitchell v Pacific Dawn Pty Ltd* [2007] QCA 74 [7].

[44] [1995] 2 AC 378 (HL) 392.

[45] John Shelton, 'Agreed Remedies: Comment' in Andrew Burrows and Edwin Peel (eds), *Commercial Remedies: Current Issues and Problems* (OUP 2003) 223.

[46] *ACCC v CG Berbatis Holdings Pty Ltd* [2003] HCA 18, (2003) 214 CLR 51 [43] (footnote omitted).

To start, an important distinction exists between 'efficiency theory' and 'economic analysis' of law.[47] 'Efficiency theory' is a private law theory that assumes that the law is, or should be, efficient. Conversely, 'economic analysis' of law is a methodological approach to legal analysis that examines the law in terms of incentives and disincentives. A scholar who supports an efficiency theory of private law will inevitably adopt an economic analysis of private law. However, a scholar may equally adopt an economic analysis to argue that the private law is *inefficient*. If so, she will use an economic analysis, but will not produce an efficiency theory of private law.

Next, under efficiency theory, the private law's *sole* aim is to promote economic efficiency.[48] An act or rule is 'efficient' if its benefits exceed its costs. In this cost-benefit evaluation, the criterion used to assess costs and benefits is 'welfare'. 'Welfare' refers to 'the satisfaction of individual preferences', and the object of these preferences include all matters upon which an individual might place value, such as goods, services, or personal fulfilment.[49] Further, efficiency theorists assume that individuals respond to incentives. Legal rules promote efficiency by acting as 'taxes' that provide incentives to individuals to act efficiently, and discourage them from acting inefficiently.

Lastly, efficiency theorists make many, central assumptions. Importantly, efficiency theory assumes that all individuals are rational actors who will try to maximise their self-interest or welfare. As such, an individual is the soundest determinant of what is in her own best interests. Accordingly, it is said that 'left to themselves, parties on equal terms will reach the agreement that maximizes the benefits to both sides'.[50]

Turning to the issue at hand, efficiency theorists adopt one basic argument purportedly to justify the Penalty Rule: preventing A from enforcing a penalty clause against B promotes efficiency. They give three principal reasons, as follows.

First, it is said that if the private law permitted parties to enforce penalty clauses, it would discourage an 'efficient breach' of contract. So, the 'efficient breach' theory holds that contract law remedies do, and should, provide an incentive to a defendant to breach his contract with a claimant where non-performance is efficient. Generally, it is efficient for a defendant to breach his obligations if that breach will lead to a situation where the defendant is

---

[47] Stephen A Smith, *Contract Theory* (OUP 2004) 109.

[48] Richard A Posner, *Economic Analysis of Law* (8th edn, Aspen Publishers 2011) [4.09].

[49] Stephen A Smith, *Contract Theory* (OUP 2004) 109.

[50] Andrew Ham, 'The Rule Against Penalties in Contract: An Economic Perspective' (1990) 17 MULR 649, 654.

factually better off, and the claimant is factually no worse off.[51] Put differently, as a result of the breach the defendant's economic welfare will increase and the claimant's economic welfare will not change, given the defendant must compensate the claimant for any deterioration in her factual position. Consequently, the outcome produced will result in an increase in society's welfare. Notably, on this view it necessarily follows that a contractual 'obligation' is in truth an option to perform, or pay compensatory damages if more efficient.[52]

An example might help. Assume that due to a change in circumstances B will make no profit from performing his contractual obligation to A, and that the direct cost to B of performing is more than the loss A will suffer as a result of B breaching his duty. Hence, B's cost of performance exceeds the amount of general compensatory damages he must pay to A. If so, then it is efficient for B to breach his obligation: it will minimise B's outlay, leaving him better off; and it will place A in the same economic position as that she would have occupied had B performed, leaving her no worse off.

Conversely, if B is obliged to pay a penalty to A upon default, B will be strongly encouraged to perform his obligation when it would be more efficient for him to breach it. In Posner's words, 'A penalty [clause] would deter efficient … breaches by making the cost of the breach to the contract breaker greater than the cost of the breach to the victim.'[53] An example:[54] D agrees to make a machine part for C for £1,000. X then offers D £1,500 for that part. C expects to make £300 profit from reselling the part. The efficient breach theory requires that D breaks the original contract with C; pays to C general compensatory damages of £300; enters an alternative contract with X; and transfers the part to X in return for £1,500. However, if the original contract contains a penalty clause requiring D to pay to C £600 if he defaults, this will prevent the efficient breach, as D will no longer be left better off by breaching his obligation to C.

Second, it is commonly argued that if Anglo-Australian private law upheld penalty clauses, it would incentivise A and B to spend resources wastefully.[55] As to A, she would spend time and money trying to induce B to breach his obligations, in the hope of securing the stipulated penal sum. In Posner's words, a penalty clause 'might induce the prospective victim

---

51 This type of efficiency is known as 'Pareto' efficiency.

52 OW Holmes, 'The Path of the Law' (1897) 10 Har L Rev 457, 462.

53 Richard A Posner, *Economic Analysis of Law* (8th edn, Aspen Publishers 2011) [4.12].

54 Andrew Burrows, *Remedies for Torts and Breach of Contract* (3rd edn, OUP 2004) 450.

55 Kenneth W Clarkson, Roger LeRoy Miller and Timothy J Muris, 'Liquidated Damages v Penalties: Sense or Nonsense?' [1978] Wis L Rev 351, 370.

to provoke a breach, since [she] would profit from it'.[56] As to B, he would devote time and money attempting to detect and prevent potential breach-inducing conduct.

Finally, it is propounded that allowing a party to enforce a penalty clause will lead to more bankruptcies and insolvencies. As B will be required to pay to A an extravagant amount if he defaults, B will more likely become financially distressed, and fail in his business. As such, Posner contends that penalty clauses 'increase the risk of bankruptcy consequent on contractual default, and by doing so increase the number and … total cost of bankruptcies …'.[57] Further, it is said that if penalty clauses were commonly used and enforced, a recession would spark a 'wave of business failures'.[58] Consequently, the number of bankruptcies and insolvencies in economic downturns would greatly increase, compared with the number that would occur if the Penalty Rule existed. Society generally would be left worse off.

The dominant academic view is that efficiency theories generally score well on the fit and coherence criterions for assessing private law theories.

As to the fit criterion, efficiency theory provides a neat explanation of several legal rules.[59] For example, it explains implied terms as filling gaps in the parties' agreement. This promotes efficiency by saving the parties from the time and expense of attempting to draft an agreement covering every potential eventuality. Also, the efficient breach theory seemingly explains contract law remedies well, for these rules appear to encourage a party not to perform his obligations *only* in cases where this is efficient. The private law apparently promotes efficient breaches in the following ways: by making general damages the primary remedy for a breach of contractual duty, not specific performance; by setting the amount of general compensatory damages at the value of the promised performance; and by only awarding specific performance where general damages are too difficult—and hence too costly—to assess.

As to the coherence criterion, this rationale has the capacity to render the private law strongly coherent. On this view, all of the private law's constituent rules (including the Penalty Rule) derive from one master, overarching concept of promoting efficiency. The private law is thus perfectly unified.

---

56 Richard A Posner, *Economic Analysis of Law* (8th edn, Aspen Publishers 2011) [4.12].

57 Richard A Posner, *Economic Analysis of Law* (8th edn, Aspen Publishers 2011) [4.12].

58 Daniel A Farber, 'Contract Law and Modern Economic Theory' (1983) 78 Nw UL Rev 303, 335.

59 Stephen A Smith, *Contract Theory* (OUP 2004) 113, 117-18.

B Evaluation

Though of long-standing, an efficiency-based explanation of the Penalty Rule is unpersuasive. The following considers the four evaluative criteria that are the most problematic: coherence, transparency, principle and practice.

I Coherence

Despite what has been said, the efficiency theory presents two problems of coherence.

For one, if all private law doctrines are means to an efficient ends, then consistency and coherence demand an 'efficient breach' theory in contract law *and* an 'efficient conversion' theory in tort law. Friedmann provides the following two examples. In the first example, D contracts to sells a machine part to C for £1,000. Instead, D then sells it to X for the sum of £1,800. C expected to make £200 profit from the sale. In the second example, C owns a machine part for which she paid £1,000. D takes the part, and sells it to X for the sum of £1,800. C expected to make £200 profit from any resale. The writer argues that, 'If the wilful contract breach is justified in the first case, then the wilful conversion is justified in the second.'[60] However, no one has ever seriously contended that efficiency-based considerations justify a converter's conduct, or tort law's remedies in response to it. This is because neither can be so justified. If that is correct, then two parts of the private law—contract law and tort law—are not consistent. Seen in this light, then, *the private law* is not even weakly coherent.

For another, courts regularly award non-compensatory remedies that deter an efficient breach, such as specific performance, injunctions and disgorgement damages. Accordingly, these remedies are clearly inefficient. Nonetheless, they are valid. This undermines the view that all private law remedies are united by the fact they promote efficiency. Further, assume for argument's sake that the Penalty Rule is properly justified on the basis that penalty clauses discourage an efficient breach. If so, in the premise Anglo-Australian contract law is inconsistent: on the one hand, it invalidates penalty clauses *merely because* they are inefficient; on the other hand, it sanctions other contract law remedies *despite* the fact they are inefficient. On this view, it follows that *contract law* is not even weakly coherent.

II Transparency

[60] Daniel Friedmann, 'The Efficient Breach Fallacy' (1989) 18 JLS 1, 4.

For three reasons, the efficiency theory fails the transparency criterion for assessing private law theories. It is not even weakly transparent. This is the theory's greatest flaw.

To start, efficiency-based reasoning does not respect the law's self-understanding.[61] As Weinrib notes, 'While economists applaud legal results that coincide with efficiency, their framework does not respect the law's characteristic features on their own terms.'[62] Put differently, efficiency theory characteristically attempts to explain private law doctrines using concepts that are foreign to legal reasoning. As Smith argues, efficiency theories are inappropriately 'external' to the private law: they do not adequately explain private law doctrines because 'they rely on concepts, language, and reasoning that are radically different from those employed by legal actors themselves'.[63] Thus, efficiency theorists employ reasoning of a completely different *kind* to the reasoning that courts typically adopt—it is not merely reasoning that is different in terminology or degree.

Take the general damages rules, for instance. For efficiency theorists, these rules merely provide *incentives*; they are tools for promoting future efficient outcomes, and deterring future inefficient behaviour. For judges, these rules provide *remedies*; they seek to cure or make better a past rights violation, and address consequential matters arising from that infringement. Courts label a damages order as a 'remedy', and it does what the label says. In short, judges do not explain what they are doing as providing incentives to parties generally to act efficiently in future. Rather, judicial reasoning is cast in terms of recognising the specific parties' rights and obligations, which necessarily arose in the past. Therefore, efficiency theory does not belong to the law's self-understanding. It is unable to explain the Penalty Rule from an appropriately 'internal' legal perspective.

Next, the transparency criterion usually refers to the *substance* of the court's reasons. However, a related aspect of transparency is the *form* of the court's reasoning processes. Some forms of legal reasoning are illegitimate in private law judicial decision-making. As such, judges cannot adopt any argument that bears one of these prohibited reasoning structures. Relevantly, the Australian High Court has held that 'top-down' reasoning is an illegitimate form of judicial reasoning in the private law. In *Roxborough v Rothmans of Pall Mall Australia Ltd*, Gummow J held that, in such reasoning 'a theory about an area of law is invented or adopted and then applied to existing decisions to make them conform to the theory and to dictate the

---

[61] Text to n 26 in ch 1.

[62] Ernest J Weinrib, *The Idea of Private Law* (Harvard UP 1995) 47.

[63] Stephen A Smith, *Contract Theory* (OUP 2004) 133.

outcome in new cases'.[64] His Honour suggested that 'top-down' reasoning transgresses common law technique. In particular, relying on Posner's work, Gummow J held that a feature of 'top-down' reasoning is that it uses a theory to accept, reject, explain away or distinguish decisions to make them conform to the theory.[65] Recalcitrant cases are resized to fit the Procrustean bed. This suggests that a 'top-down' theory is *prescriptive* in nature.

As discussed in Chapter 1, a prescriptive account of a legal area outlines what the law ought to be. Clearly, courts cannot adopt a purely prescriptive theory of what the private law ought to be. It would result in judges deciding cases according to preconceived views, not according to the evidence, thus transgressing the rule of law. Also, if judges could legitimately adopt a purely prescriptive approach, nothing would be left of the doctrine of precedent.[66]

Efficiency theories do not contend that, as best explained, the underlying basis of all private law rules is that they function to promote economic efficiency. Rather, they argue that private law doctrines *should* promote economic efficiency. In Weinrib's words, efficiency theorists are 'not so much concerned with understanding private law as with assessing the degree to which its rules coincide with what efficiency demands'.[67] Posner confirms this. The writer notes that he is 'associated with several top-down theories', as follows:

> One, which is primarily … [descriptive], is that the common law is best understood on the 'as if' assumption that judges try to maximize the wealth of society. Another, primarily normative, is that judges should interpret … statutes to make them conform to the dictates of wealth maximization.[68]

The second theory is clearly prescriptive. Despite Posner's contrary statement, so is the first. The language is telling: 'as if' encompasses a fiction. The fiction is that judges seek to maximise society's wealth. This implies, in turn, that judges do so because the private law's basic purpose is to promote efficiency. That fiction encapsulates a normative judgment that promoting efficiency is precisely what judges should in fact be doing. In short, efficiency theory adopts a purely proscriptive, 'top-down' reasoning structure that judges are not free to

---

[64] [2001] HCA 68, (2001) 208 CLR 516 [73]. See also *Lumbers v W Cook Builders Pty Ltd (In Liq)* [2008] HCA 27, (2008) 232 CLR 635 [77].

[65] *Roxborough v Rothmans of Pall Mall Australia Ltd* [2001] HCA 68, (2001) 208 CLR 516 [73], citing Richard A Posner, 'Legal Reasoning From the Top Down and From the Bottom Up' (1992) 59 U Chi L Rev 433, 433.

[66] See Carmine Conte, 'From Only the 'Bottom-up'? Legitimate Forms of Judicial Reasoning in Private Law' (2015) 35 OJLS (forthcoming).

[67] Ernest J Weinrib, *The Idea of Private Law* (Harvard UP 1995) 48.

[68] Richard A Posner, 'Legal Reasoning From the Top Down and From the Bottom Up' (1992) 59 U Chi L Rev 433, 433.

adopt. Therefore, efficiency theory does not belong to orthodox, permissible judicial methods. Consequently, it is unable to justify the Penalty Rule adopting a legitimate judicial methodology.

Lastly, the Australian High Court has consistently dismissed key aspects of the efficiency theory in the general damages context. In *Coulls v Bagot's Executor and Trustee Co Ltd*, Windeyer J held that it was 'a faulty analysis of legal obligations to say that the law treats a promisor as having a right to elect either to perform his promise or to pay damages'.[69] Recently, in *Tabcorp Holdings Ltd v Bowen Investments Pty Ltd*, the High Court noted that several remedies exist that deter an efficient breach, such as specific performance and injunctions; and that the efficient breach theory is flawed in that it 'underrates the extent to which those remedies are available'.[70] The efficient breach notion, then, is anathema to the judges on Australia's highest court. It is therefore anathema to a transparent theory.

III Principle

The efficiency theory raises four issues of principle.

First, a fundamental premise of efficiency theory completely undermines any efficiency-based justification of the Penalty Rule. As stated, efficiency theory assumes that an individual is the soundest determinant of what is in her own best interests. Accordingly, the private law should not alter the parties' voluntary, express agreement. In Posner's view, voluntarily negotiated penalty clauses 'would not be included in contracts … unless the parties expected the gains to outweigh the costs …'.[71] It follows that the law should allow the parties to determine their own best interests by including an enforceable penalty clause in their contract, if they so wish. Properly analysed, then, efficiency theory supports *excising* the rule from Anglo-Australian private law.

Second, in truth, permitting A to enforce a penalty clause against B will not discourage an efficient breach. A rational, commercial B will 'negotiate out' of paying a penalty to A, thus allowing an efficient breach to occur.[72] For example, B may offer a settlement sum to A comprising the total amount that A would have made from B's performance, plus a proportion of the additional gain that B will make from performing a more profitable contract with C.

---

[69] (1967) 119 CLR 460 (HCA) 504.

[70] [2009] HCA 8, (2009) 236 CLR 272 [13].

[71] Richard A Posner, *Economic Analysis of Law* (8th edn, Aspen Publishers 2011) [4.12].

[72] Charles J Goetz and Robert E Scott, 'Liquidated Damages, Penalties and the Just Compensation Principle: Some Notes on an Enforcement Model and a Theory of Efficient Breach' (1977) 77 Colum L Rev 554, 568.

This is the Coase Theorem, simply applied. The Coase Theorem holds that, if no transaction costs exist, legal rules are irrelevant because the parties—as rational maximises of value—will bargain around those rules to generate the most efficient outcome.[73] Given A and B are already known to each other, the transaction costs are unlikely to be sufficiently high to prevent a negotiated release. In short, private bargaining will achieve an efficient breach irrespective of the legal rule.

Third, moving to the breach inducement argument, at best, it displays little explanatory force; at worst, it is inherently flawed.

As to the former, the logic of the breach inducement argument suggests that the Penalty Rule should only apply where A can 'covertly increase the probability of breach and when [she] might have incentive to do so'.[74] However, presently this does not form part of the jurisdictional hurdle that B must surmount. Further, it is said that courts should not enforce an agreed damages clause that is 'reasonable *ex ante*' if circumstances change after the contract date rendering it 'unreasonable *ex post*', because the changed circumstances 'create an incentive both to induce wasteful breach and to monitor for breach inducement'.[75] However, the tests of validity are applied as at the contract date. Accordingly, in the premise, the agreed damages clause is unquestionably enforceable.

As to the latter, if Anglo-Australian courts upheld penalty clauses, the 'prevention principle' would still exist.[76] This principle holds that X cannot insist on Y's performance of a contractual obligation if X is the cause of Y's non-performance. If the prevention principle applies, X is prevented from enforcing against Y her contractual rights that arise upon Y's non-performance.[77] The principle commonly applies in the construction contract context. For instance, if a principal prevents a contractor from completing within the fixed period, the principal will lose its right to claim liquidated compensatory damages from the contractor. So, if courts *prima facie* enforced penalty clauses, pursuant to the prevention principle a court would not do so where A caused B to breach his obligation under the clause. In short, if A 'provoked' B's breach, she would be disentitled from enforcing her right to the stipulated penal

[73] RH Coase, 'The Problem of Social Cost' (1960) 3 JLE 1.

[74] Kenneth W Clarkson, Roger LeRoy Miller and Timothy J Muris, 'Liquidated Damages v Penalties: Sense or Nonsense?' [1978] Wis L Rev 351, 375.

[75] Kenneth W Clarkson, Roger LeRoy Miller and Timothy J Muris, 'Liquidated Damages v Penalties: Sense or Nonsense?' [1978] Wis L Rev 351, 375.

[76] *Spiers Earthworks Pty Ltd v Landtec Projects Corporation Pty Ltd (No 2)* [2012] WASCA 53*,* (2012) 287 ALR 360 [47]-[49].

[77] X's act of prevention may constitute a breach of contract itself, but this need not necessarily be the case.

sum. It follows that, even if the Penalty Rule did not exist, A would have no incentive to spend her resources trying to induce B to breach his obligations.

Finally, Posner's argument regarding increased insolvencies overreaches. It overreaches because it suggests that courts are entitled to make decisions based on consequentialist issues, when such concerns properly lie beyond a court's institutional competence. In Beever's words, 'judges are not in a position to make these decisions, both because they lack the necessary information and because they usually lack the training to deal in an informed matter with the information if they had it'.[78] Notably, in this context, the relevant 'information' is said partly to constitute the potential for a bushfire of business collapses if penalty clauses are routinely enforced. This consequence is not self-evident as a matter of common sense. Nor is any evidence cited in support of it. It is pure conjecture.

IV Practice

At least two practical concerns exist here, as follows.

One concern is Posner's practical argument in support of an efficiency-based Penalty Rule. The writer contends that if the courts automatically enforced penalty clauses it would lead to more litigation, and more litigation costs. Thus, it is said that if penalty clauses were enforceable, A would likely 'hold out for something very near the penalty [sum]', and as a result B would probably 'decide he might as well let [A] sue'.[79] That is simply incorrect. No competent practitioner would advise this. If the legal outcome is clear, there is little reason for the parties to litigate. In fact, automatic enforcement would result either in the same amount of litigation, or more likely in a decrease in litigation. Where litigation in fact ensues, it would cost the parties and society less, as courts would no longer need to spend time applying the rule.

A larger concern is this. Practising lawyers are 'understandably uncomfortable with an explanation of the law that is so at odds with how they understand what they are doing'.[80] If the efficiency theory is correct, then practitioners have understood the Penalty Rule, and provided advice to clients, in misleading terms. This can only discredit the law. Client hostility would be understandable.

---

[78] Allan Beever, *Rediscovering the Law of Negligence* (Hart Publishing 2007) 172.

[79] Richard A Posner, *Economic Analysis of Law* (8th edn, Aspen Publishers 2011) [4.12].

[80] Stephen A Smith, *Contract Theory* (OUP 2004) 134.

V Conclusion

In conclusion, the efficiency theory displays several intractable problems. Most importantly, it fails the transparency criterion for assessing private law theories. The externality of the efficiency-based explanation of the rule leaves a gaping hole in its explanatory power.

An important point should not be missed. There is no objection to *assessing* the Penalty Rule according to external criteria, such as efficiency, morality or gender equality. But it cannot be *understood*, and hence *justified*, by reference to such criteria. If we are searching for the rule's justificatory explanation, we cannot turn away from the law to consider something else. An extrinsic criterion like efficiency is, literally, a distraction.

It follows that the efficiency theory cannot justify the Penalty Rule. In fact, Posner admits this. The writer concedes that the rule remains 'a major *unexplained* puzzle' in the efficiency theory of the private law.[81] The efficiency rationale, then, is a failure.

3 Preventing 'overcompensation'

A Thesis

Some argue that the 'true source' of the Penalty Rule lies in 'the prevention of overcompensation'.[82] This proposed theoretical justification is orthodoxy.

The preventing 'overcompensation' rationale sees the rule as *prescriptive* in nature. It requires the parties appropriately to mimic general compensatory damages in any agreed damages clause. Any agreed damages clause that entitles A to recover from B a sum that exceeds the amount of general compensatory damages that a court would award is penal, and unenforceable. The theory is so obvious that commonly it is simply assumed. It is never set out in any detail. Its underlying reasoning can be summarised in five steps.

First, as traditionally understood, in awarding general damages for breach of contract, Anglo-Australian courts apply the 'compensatory principle'. This principle requires that a court award a sum of money that, as far as money can be so, is equivalent to the claimant's pecuniary or non-pecuniary loss that the defendant's breach has caused. Thus, the courts award general damages for breach of contract to compensate the claimant for her loss. For

81 Richard A Posner, 'Some Uses and Abuses of Economics in Law' (1979) 46 U Chi L Rev 281, 290 (emphasis added).

82 Kanaga Dharmananda and Leon Firios, 'Penalties Arising Without Breach: The Australian Apogee of Orthodoxy' [2013] LMCLQ 145, 150.

example, in *Ng v Ashley King (Developments) Ltd*, Lewison J held that, 'The overriding principle in any assessment of damages for breach of contract is to compensate the injured party rather than to punish the contract breaker'.[83]

Second, the general damages and agreed damages regimes are interconnected. In the language of Chapter 5, the private law adopts an associated view. The agreed damages regime is merely an *analogue* of the general damages regime. Consequently, under any agreed damages clause, A may only recover from B an amount that suitably reflects the amount that a similarly situated claimant would recover from a defendant under a general law remedy. As Burrows argues, the validity of an agreed damages clause will 'be assessed in line with the judicial remedies available. The further the parties stray from what the courts would themselves award, the more likely it is that the agreed remedy will be regarded as penal …'.[84]

Third, it follows that, in agreeing damages, the parties must apply the compensatory principle, and must imitate a judicial compensatory remedy. Accordingly, in assessing the relevant amount, an agreed damages clause must appropriately account for the general law principles for calculating general compensatory damages. In theory, this should produce an agreed compensatory damages sum that properly mirrors the quantum of a general compensatory damages award. If not, the agreed damages clause is penal. McKendrick expresses the idea thus: 'it is for the courts to decide the compensation which is payable upon a breach of contract and … it is not for the parties to set compensation at a level higher …'.[85] Therefore, there is only one type of liquidated damages, and that is liquidated *compensatory* damages. The two concepts are synonymous.

Fourth, as the name of this rationale suggests, the Penalty Rule is concerned with only one response: compensation. The rule is 'simply a manifestation of the law's fascination with compensation as the principal remedy for breach of contract…'.[86] The corollary of this obsession with compensation is that Anglo-Australian law prohibits 'overcompensation'. Thus, Bryan has said that the rule is 'grounded in the policy of preventing overcompensation for breach of contract'.[87] Accordingly, it is said that the private law prohibits 'judges from

---

[83] [2010] EWHC 456 (Ch), [2011] Ch 115 [51(i)]. See similarly *Hospitality Group Pty Ltd v Australian Rugby Union Ltd* [2001] FCA 1040, (2001) 110 FCR 157 [159].

[84] Andrew Burrows, *Remedies for Torts and Breach of Contract* (3rd edn, OUP 2004) 445.

[85] Ewan McKendrick, *Contract Law* (10th edn, Palgrave Macmillan 2013) [21.8].

[86] JW Carter and Michael Tilbury, 'Remedial Choice and Contract Drafting' (1998) 13 JCL 5, 11.

[87] Michael Bryan, 'Equitable Relief from Forfeiture: Performance or Restitution?' in Charles Rickett (ed), *Justifying Private Law Remedies* (Hart Publishing 2008) 379. See similarly Solene Rowan, 'For the Recognition of Remedial Terms Agreed Inter Partes' (2010) 126 LQR 448, 461.

making awards of damages in excess of the strict requirements of compensation, but also it prevents parties from contracting around this rule'.[88]

Finally, the distinction between primary and secondary rights is critical. On a rights-based analysis, where a defendant violates a claimant's primary claim right, a secondary claim right arises. So, where Y breaches his contractual duty to X, X obtains a secondary right against Y for general damages. Similarly, where the contract contains an agreed damages clause, upon breach A obtains a secondary right against B for agreed damages. On the preventing 'overcompensation' theory, judges widely respect the parties' primary rights, but closely scrutinise secondary rights because they are 'the preserve of the courts'.[89]

As the rule's traditional rationale, the preventing 'overcompensation' theory convincingly explains the outcome of most cases, and the path the relevant court adopted to reach the particular outcome. Accordingly, the theory scores highly on the fit and transparency criterions.

Beginning with the fit criterion, the preventing 'overcompensation' theory displays a tight fit with various aspects of the case law. Two are noteworthy.

On the one hand, this rationale is consistent with fundamental aspects of the jurisdictional test, the test of validity, and the rules regarding consequential relief.

To start, this putative rationale rationalises the breach requirement, and why the paradigm scenario[90] involves a breach. Simply put, if a wrong is a prerequisite for a court to award general damages, so too for the parties to provide for liquidated damages. Hence, only where B has committed a breach may the court review the clause to ensure that it does not overstep its proper boundaries.[91]

Next, the preventing 'overcompensation' theory is supported by the tests of validity enunciated in *Dunlop*. The various tests use the terminology of 'damage', 'loss' and 'compensation'.[92] Further, later case law has emphasised the Estimation Test and the Extravagance Test. These tests require a comparison with A's anticipated loss.

---

[88] Hugh Collins, 'Fairness in Agreed Remedies' in Chris Willett (ed), *Aspects of Fairness in Contract* (Blackstone 1996) 105.

[89] Mindy Chen-Wishart, 'Controlling the Power to Agree Damages' in Peter Birks (ed), *Wrongs and Remedies in the Twenty-First Century* (Clarendon Press 1996) 273.

[90] Text following n 20 in ch 1.

[91] DW Greig and JLR Davis, *The Law of Contract* (Lawbook Company 1987) 1465.

[92] *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd* [1915] AC 79 (HL) 86-87.

Lastly, this proposed rationale is consistent with the fact that the rule requires A partially to enforce a penalty clause against B, and not to bring a general damages claim. Although the parties 'step into the shoes' of the court to award damages, they undertake this judicial-like function by means of *the agreement*. This necessarily requires that a court partially enforce the valid part of the agreement, and not enforce the invalid 'overcompensatory' part.

On the other hand, the preventing 'overcompensation' theory provides an economical and powerful explanation of the outcomes in several cases. Consider two examples.

An Australian example is the 'central' case of *AMEV-UDC Finance Ltd v Austin*.[93] As Chapter 3 discussed, the Australian High Court held that the relevant clause was penal because it failed to provide for two factors: first, a rebate for the accelerated payment of future instalments; and, second, a credit for the anticipated sale proceeds if they exceeded the residual value of the goods. So, in other words, the agreed damages clause did not allow for A to recover from B the total anticipated rental less the excess of the anticipated resale price over the residual value of the goods, nor did it discount for early receipt. On this rationale, the reason why the clause is invalid is simple.[94] It failed properly to take account of the mitigation doctrine (which applied given a resale was sufficiently certain); and of the fact that general damages must be appropriately discounted to reflect the net present value of money.

The English example is the 'central' case of *Makdessi v Cavendish Square Holdings BV*.[95] As Chapter 5 noted, the English Court of Appeal endorsed the Commercial Justification Test. It relied upon that test in deciding that the two agreed damages clauses it was considering were penal. This thesis has demonstrated that the Commercial Justification Test is irreparably flawed. The Court's primary reasoning, therefore, is suspect. However, if one adopts a preventing 'overcompensation' theory, the reason for the decision is straightforward.[96] So, Cavendish was seeking to enforce two agreed damages clauses against Makdessi. As at the contract date, the anticipated loss that Cavendish would suffer if Makdessi breached his obligations under either clause was a loss of the value of its shareholding in another company. Such loss is a 'reflective' loss that is legally irrecoverable.[97] Accordingly, Cavendish's anticipated legally-recoverable loss was zero. It followed that the two agreed damages sums

[93] (1986) 162 CLR 170 (HCA).

[94] cf JW Carter and Elisabeth Peden, 'A Good Faith Perspective on Liquidated Damages' (2007) 23 JCL 157, 174.

[95] [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125.

[96] See similarly Edwin Peel, 'Unjustified Penalties or an Unjustified Rule Against Penalties?' (2014) 130 LQR 365, 367 cf Janet O'Sullivan, 'Lost On Penalties' (2014) 73 CLJ 480, 483.

[97] *Johnson v Gore Wood and Co* [2002] 2 AC 1 (HL) 66.

were clearly penal. As the Court recognised, 'If the [legally] recoverable figure is zero, any estimate other than zero is excessive. *A fortiori* a figure in millions is extravagant in amount …'.[98]

Turning to the transparency criterion, the preventing 'overcompensation' rationale is strongly transparent. Two examples follow.

For one, many judges have taken an associated view, and held that the agreed damages regime is an analogue of the general damages regime. For instance, in *Citicorp Australia Ltd v Hendry*, Mahoney JA held that the Penalty Rule is based on a rationale of 'limiting recovery to the damages recoverable at law'.[99]

For another, several judges have emphasised the rule's concern with 'overcompensation', and its apparent focus on secondary rights. So, in *Photo Production Ltd v Securicor Transport Ltd*, Lord Diplock classically stated that an agreement:

> must not impose upon the breaker of a primary obligation a general secondary obligation to pay to the other party a sum of money that is manifestly intended to be in excess of the amount which would fully compensate the other party for the loss sustained by him in consequence of the breach of the primary obligation.[100]

B Evaluation

The preventing 'overcompensation' theory is conventional. Accordingly, over the years, it has come under fire. This section examines the critiques of the rationale under the following headings: coherence, fit and principle. Notably, the objections so far put against the theory do not withstand close scrutiny. That said, some previously unarticulated problems with it do exist.

I Coherence

Some scholars argue that the alleged concern to prevent 'overcompensation' is a selective one, for the private law allows a degree of 'overcompensation' in other contexts. If this assertion is correct, then it follows that any preventing 'overcompensation' theory of the Penalty Rule results in a private law that is internally inconsistent. In other words, it is not even weakly

[98] *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [107] (original italics).

[99] (1985) 4 NSWLR 1 (NSWCA) 30. See similarly *Robophone Facilities Ltd v Blank* [1966] 1 WLR 1428 (CA) 1447.

[100] [1980] AC 827 (HL) 850.

coherent. There are two principal areas where Anglo-Australian private law is said to sanction 'overcompensation', outlined below. Neither is convincing.

On the one hand, Cunnington argues that the private law allows 'overcompensation' where a promisor breaches an 'entire' obligation, for 'the promisee ends up with both the price of the works (because he does not have to pay for it) and the [value of the] partially completed works'.[101] The writer cites *Sumpter v Hedges* in support.[102] A builder contracted with an owner to build two houses and stables on the owner's land. The builder completed much of the construction work. Then, he experienced financial difficulties and was unable to finish. The builder sought an order for reasonable remuneration, totalling the value of the services rendered. The English Court of Appeal rejected the claim. However, the result in this case has nothing to do with compensation. Properly analysed, the builder was a contract-breaker seeking restitution based on the unjust factor of failure of basis. The effect of the Court rejecting the claim was that the owner was not required to make restitution of the value of the partially completed works. Whether that result is justified or not cannot be pursed here. For present purposes, the important point is that the causative event in play was unjust enrichment, and the response was *restitution*. The case had nothing to say about *compensation*. These two responses are conceptually distinct.

On the other hand, Coote contends that the law of general damages permits 'overcompensation' because, upon a buyer's breach, a seller of shares is entitled to the difference between the contract price and market price, 'notwithstanding resale at a higher price'.[103] However, the result in such a case has nothing to do with compensation. Properly interpreted, that case involves damages that substitute for the value of an infringed right, where the value of the right is assessed by a difference in value calculation. That is, in the case of non-acceptance under a sale of goods contract, the value of a contractual right to sell goods at a particular price is the difference between the net contract price and the market value of the proprietary right which is the subject of sale, as at the breach date. Further, as Chapter 1 explained, compensatory damages are assessed at the judgment date. In assessing the claimant's consequential loss, a court will take into account events that have occurred after the wrong that have increased or decreased the loss that she would normally suffer. Conversely, substitutive damages are assessed at the time of infringement, such that 'Subsequent events are

---

[101] Ralph Cunnington, 'Should Punitive Damages be Part of the Judicial Arsenal in Contract Cases?' (2006) 26 LS 369, 382 (footnote omitted).

[102] [1898] 1 QB 673 (CA).

[103] Brian Coote, 'Contract Damages, Ruxley, and the Performance Interest' (1997) 56 CLJ 537, 548-49.

irrelevant as the court's task is not to calculate what [incurred] loss has been suffered.'[104] The non-acceptance cases, therefore, are irrelevant to any argument concerning the compensation response.

II Fit

Chen-Wishart asserts one overarching argument of fit: the Penalty Rule tests of validity produce results that are not co-extensive with the outcomes that the compensatory principle generates.[105] In summary, a general compensatory damages award implements the compensatory principle. Further, the compensatory principle underlies the preventing 'overcompensation' theory. However, a court may hold that an agreed damages clause is valid despite the fact that the agreed damages sum is less than, or more than, the amount of a general compensatory damages award. Also, a court may hold that such a clause is penal even though the stipulated sum precisely equals the amount of a general compensatory damages award. It seemingly follows that the preventing 'overcompensation' theory is problematic, because the tests of validity sanction outcomes that contradict the compensatory principle said to underlie the theory.

This contention is superficially attractive. However, upon close inspection, it is misconceived. In what follows, this thesis examines separately the situations where the agreed damages sum is equal to, more than, and less than a general compensatory damages amount awardable for the same breach.

(1) Equal to

Chen-Wishart argues that a general compensatory damages award 'accords with the compensatory principle'.[106] Consequently, the writer contends that the tests of validity are unfaithful to that principle because an agreed damages clause that sets an amount equal to a general compensatory damages award may be held penal. In support, she proffers the example of an agreed damages sum that the Greater Sum Test invalidates.

---

[104] Robert Stevens, 'Damages and the Right to Performance: a Golden Victory or Not?' in Jason W Neyers, Richard Bronaugh and Stephen GA Pitel (eds), *Exploring Contract Law* (Hart Publishing 2009) 182.

[105] Mindy Chen-Wishart, 'Controlling the Power to Agree Damages' in Peter Birks (ed), *Wrongs and Remedies in the Twenty-First Century* (Clarendon Press 1996) 277-78.

[106] Mindy Chen-Wishart, 'Controlling the Power to Agree Damages' in Peter Birks (ed), *Wrongs and Remedies in the Twenty-First Century* (Clarendon Press 1996) 277.

As Chapter 5 demonstrated, the Greater Sum Test no longer works. Under the Penalty Rule as best interpreted, this test is interpreted away. However, one (obsolete) example of the point does not establish its validity. The point is in fact incorrect, as a general proposition. In principle, a test of validity that invalidates an agreed compensatory damages amount that is the same as a general compensatory damages amount that a court would award for the same breach will *not* violate the compensatory principle. This is because liquidated damages are only a general damages *analogue*. They are not a *replica*.

To explain, two differences exist between liquidated and general damages. On the one hand, it is the parties, and not the court, assessing the liquidated damages payable. The parties effectively 'step into the shoes' of the court for the purpose of assessing the damages payable for B's breach. On the other hand, the assessment of a liquidated damages sum is conducted at the contract date, and not at a later date. Liquidated damages and general damages, therefore, are cognate concepts. But the former is not a facsimile of the latter.

So, the parties assess A's *anticipated* loss *ex ante*, whereas the courts assess A's *incurred* loss *ex post*. Given this, the process of entering into, or reviewing, an agreed compensatory damages clause requires a conceptually different approach—'a different methodology'[107]—to that employed in calculating a general compensatory damages award. Courts assess general compensatory damages at the judgment date, taking into account the factual detriment that A has suffered. Conversely, in assessing agreed damages, the parties' inquiry necessarily takes place at the contract date. Similarly, when reviewing an agreed damages clause, a judge must notionally measure the anticipated loss as at the contract date. The judge must situate herself at a single point in time, and assess the anticipated loss exclusively from that perspective. Consequently, a judge properly must ignore all events that occurred after the relevant assessment date, as the court's task is not to calculate the loss that A has *in fact* incurred. Further, the prospective and retrospective processes will produce a different result. Judged as at the contract date, assessing matters relating to a potential breach involves a 'necessarily speculative' prospective assessment.[108]

Thus, the tests of validity *do* require that an agreed compensatory damages clause implement the compensatory principle. However, due to the prospective nature of the assessment exercise, that principle must be applied contextually as *at the contract date*, not as at the judgment date. The correct question is: as at the contract date, what is the amount of

---

[107] *State of Tasmania v Leighton Contractors Pty Ltd* [2005] TASSC 133, (2005) 15 Tas R 243 [31(5)].

[108] *Spiers Earthworks Pty Ltd v Landtec Projects Corporation Pty Ltd (No 2)* [2012] WASCA 53, (2012) 287 ALR 360 [22].

future loss falling within the compensatory principle that A might incur? What matters is the relationship between the agreed compensatory damages sum, assessed *ex ante*, and a general compensatory damages award, if it were assessed *ex ante*. The agreed damages sum must not be extravagant compared with the court's evaluation.[109] The relationship between the agreed compensatory damages sum and a general compensatory damages award, assessed *ex post*, is neither here nor there. The private law does not expect, let alone require, equivalence.

In short, Chen-Wishart's critique fundamentally misunderstands how the compensatory principle operates in the agreed damages context.

(2) More than

An agreed damages amount that is more than a general compensatory damages sum is nevertheless valid. As such, several writers have argued that the tests of validity are inconsistent with the compensatory principle. For instance, Chen-Wishart contends that the tests of validity do not respect the compensatory principle because 'an agreed damages clause which exceeds the court assessed measure may … be enforceable'.[110] According to the writer, the reason for this inconsistent result is that these tests disregard the general law damages rules that render certain heads of loss legally irrecoverable.

One must first deal with the primary allegation. As discussed, the distinct assessment dates result in discrepancy between a liquidated compensatory damages sum and a court's compensatory award. But, in principle, this is unproblematic. The fact that, for a relevant breach, the liquidated compensatory damages exceed the general compensatory damages a court would award does not *of itself* prove that the tests of validity contravene the compensatory principle. In almost every case, what the compensatory principle requires when applied prospectively will differ to what it requires when applied retrospectively.

Many writers have failed to appreciate this point. It has led them to make statements that have muddied the waters. For example, Varghese has argued that the Penalty Rule does not display 'any inherent objection to one party being over-compensated …'.[111] This statement is ambiguous. If the stated 'overcompensation' is assessed *ex ante*, then on the orthodox view the statement is wrong. Preventing 'overcompensation'—assessed prospectively—appears to

[109] *Robophone Facilities Ltd v Blank* [1966] 1 WLR 1428 (CA) 1447.

[110] Mindy Chen-Wishart, 'Controlling the Power to Agree Damages' in Peter Birks (ed), *Wrongs and Remedies in the Twenty-First Century* (Clarendon Press 1996) 278. See similarly JW Carter and Elisabeth Peden, 'A Good Faith Perspective on Liquidated Damages' (2007) 23 JCL 157, 162.

[111] Rosslyn T Varghese, 'Penalties and Agreed Damages' (1992) 8 BCLRS 270, 271.

be the doctrine's precise basis. Conversely, if the passage refers to 'overcompensation' assessed *ex post*, then the statement it is true, but misleading. What has in fact happened is irrelevant. To focus discourse upon subsequent events is likely to lead to confusion and error.

Assuming for argument's sake that the primary allegation is correct, Chen-Wishart's supporting reasoning is fallacious. The tests of validity do not overlook the general law damages rules that exclude recovery for certain heads of loss. One example will do.

Beale suggests that A may recover from B for a non-pecuniary loss of business reputation under an agreed compensatory damages clause, even though the general law would exclude recovery.[112] A case in point is seemingly *Tall v Ryland*.[113] The plaintiff and defendant were fishmongers who conducted business in neighbouring shops. The plaintiff gave the defendant a bond of £20 conditioned to behave himself civilly towards the defendant, and not to disparage the defendant's goods. Later, the plaintiff said to one of the defendant's customers that the defendant's fish stunk. The defendant lost the customer. The plaintiff alleged that the bond was penal; the defendant contended that it was valid. Lord Keeper found in the defendant's favour. However, his Lordship suggested that the outcome may well have differed if the facts had involved a bond of '£100 or the like'.[114]

At first blush, it seems that the defendant recovered for the anticipated non-pecuniary loss of reputation resulting from the plaintiff disparaging his goods. It necessarily follows that, in applying the test of validity, the Court disregarded the general law rule that a non-pecuniary loss of reputation is irrecoverable.[115] Another interpretation exists, however. On this reading, the bond was valid because it properly estimated the anticipated *pecuniary* loss that would flow from the injury to the defendant's business reputation that the plaintiff's breach caused. A future consequential loss in sales of £20 was quite possible. By contrast, an estimated anticipated loss in custom of £100 was clearly excessive. As conventionally understood, at general law a pecuniary loss consequent upon an injury to one's business reputation is recoverable.[116]

---

112 Hugh Beale, *Remedies for Breach of Contract* (Sweet and Maxwell 1980) 57.

113 (1670) 1 Chan Cas 183, 22 ER 753.

114 *Tall v Ryland* (1670) 1 Chan Cas 183, 184; 22 ER 753, 753.

115 *Addis v Gramophone* [1909] AC 488 (HL) 496.

116 *Malik v Bank of Credit and Commerce International SA* [1998] AC 20 (HL) 40, 52; *Carter on Contract* [42-120] (service 37, April 2014).

In summary, liquidated compensatory damages that exceed a judicial compensatory award do not violate the compensatory principle. They merely involve a contextual application of that principle.

(3) Less than

As Chapter 5 illustrated, an agreed damages amount that is much less than a general compensatory damages award is nonetheless valid. However, why an agreed compensatory damages clause that underestimates A's anticipated loss is valid remains a 'controversial issue'.[117] The issue is 'controversial' because the compensatory principle—which underlies the preventing 'overcompensation' theory—seemingly requires that a derisory agreed compensatory damages clause be invalid. Thus, Chen-Wishart asserts that the tests of validity are not coterminous with the compensatory principle because 'agreed damages clauses which fall short, even far short, of compensatory damages are enforceable'.[118]

In fact, the compensatory principle is perfectly consistent with the outcome in *Cellulose Acetate Silk Co Ltd v Widnes Foundry (1925) Ltd*[119] and its progeny. To understand properly the law, one must focus on the different rights operating in this context. A rights-based analysis uncontroversially explains this issue.

Thus, a derisory liquidated compensatory damages clause does not violate the compensatory principle because in such cases it is A's *right of action*, not A's claim right, that is truncated. Where the parties' contract contains a derisory agreed compensatory damages clause, the clause is best understood in the following terms: the parties set A's *claim right* to an agreed damages sum at an amount that is equal to the general damages amount A would recover, prospectively assessed; but they simultaneously fix A's *right of action* for the agreed damages (and concomitantly B's liability) at a particular, minuscule amount. The truncated right of action prevents A from suing upon her underlying claim right in court. As such, A's claim right to compensation appropriately mirrors the amount a court would award, applying the compensatory principle as at the contract date. It follows that the clause contextually implements the principle. Further, the clause will satisfy the tests of validity, for the value of A's *claim right* (and concomitantly B's obligation) is not extravagant when compared with amount of A's anticipated loss. In fact, the two figures are precisely the same. In short, B's

---

117 Andrew Burrows (ed), *English Private Law* (3rd edn, OUP 2013) [21.180].

118 Mindy Chen-Wishart, 'Controlling the Power to Agree Damages' in Peter Birks (ed), *Wrongs and Remedies in the Twenty-First Century* (Clarendon Press 1996) 277-78 (footnote omitted).

119 [1933] AC 20 (HL).

duty to pay the higher sum remains; it is merely that A cannot completely sue upon that duty, as A's right of action (or power to sue) is cut back.

Judicial support for this analysis exists. Thus, in *Multiplex Constructions Pty Ltd v Abgarus Pty Ltd*, Cole J held that the Penalty Rule permits the parties 'to agree to *accept* and pay less than a full entitlement as damages for [a] specified breach of contract'.[120] An agreement to *accept* less suggests a setting of *liability*. The underlying *claim right* to the entire amount necessarily remains.

An example may help. Assume that A and B include an agreed compensatory damages clause in their contract which states that, upon B's breach, B is obligated to pay to A the sum of £500. As at the contract date, an appropriate estimate of A's anticipated loss is £10,000. If so, A's right of action is expressly truncated to £500. She may only sue for that amount. Conversely, A's claim right is impliedly fixed at £10,000. The clause passes the test of validity because the agreed damages sum (being the value of the claim right) is clearly not excessive when compared to A's prospective losses; in fact, the two figures are equivalent.[121]

Three important points follow.

First, in *J-Corp Pty Ltd v Mladenis*, the West Australian Court of Appeal held that a derisory liquidated compensatory damages clause is 'akin' to a limitation clause.[122] The kinship rests on the fact that both clauses 'cut back' on the defendant's liability, and concomitantly restrict the claimant's right of action. That, however, is where the similarity ends.[123]

Second, although a derisory agreed compensatory damages clause is in some aspects unusual, anatomically it remains an agreed damages clause. The clause bears the structure of, and displays the same features as, an agreed damages clause. Thus, the clause is an 'agreed damages clause' as defined in this thesis: it is a provision whereby the parties agree on the damages that B is obliged to pay to A if he breaches a term. However, the clause also decides

---

[120] (1992) 33 NSWLR 504 (NSWSC) 519 (emphasis added).

[121] This argument may save the Estimation Test from the fit-based critique put against it in Chapter 5. That is, on this interpretation, A's claim right to an agreed damages sum is clearly a substantially accurate pre-estimate of her loss. However, the Penalty Rule is nonetheless best interpreted as applying a sole test of validity, being the Extravagance Test. At least three reasons exist. First, this view is more transparent, as it is closer to the judicial reasoning in the modern Penalty Rule cases. Second, this interpretation is more principled, as the preferable nomenclature reduces ambiguity: 'genuine' can be interpreted in several different ways, 'extravagant' cannot. Finally, as Chapter 9 shows, this view is more theoretically sound, as it better encapsulates the rule's true juristic basis.

[122] [2009] WASCA 157, (2010) 26 BCL 106 [36]. See similarly Harvey McGregor, *McGregor on Damages* (19th edn, Sweet and Maxwell 2014) [15-091].

[123] Text accompanying and following n 41 in ch 5.

the damages for which A may *sue* B upon breach. Further, the clause clearly displays the two central features of an agreed damages clause, of supplanting and quantification.[124] As to the latter, the clause in fact quantifies the value of each of the two rights in play. Moreover, as with any liquidated damages clause, A may recover from B the amount of derisory liquidated compensatory damages without proving loss.

Finally, this explanation is entirely consistent with corrective justice theory. Chapter 8 will examine the substance of this theory. For now, suffice it to say that pursuant to the corrective justice norm, it is unjustifiable to undercompensate a claimant. As Beever explains, 'If a wrongdoer wrongly causes loss to a wrong-sufferer worth £100, then correcting this wrong requires the wrongdoer to pay the wrong-sufferer £100, *no less* and no more.'[125] Importantly, on this view, A's claim right to an appropriate compensatory sum remains extant. Consequently, B's obligation to pay remains in place, even though B's liability is restricted. Therefore, no unacceptable 'undercompensation' exists.

III Principle

(1) The old

In critiquing the Penalty Rule's conventional rationale, Chen-Wishart raises two arguments of principle. Neither is persuasive.

For one, the writer argues that the distinction between primary and secondary rights and obligations lacks substantive merit. In particular, Chen-Wishart states that 'no sensible distinction' exists between a promise to pay at an overvalue and a promise to pay damages which overvalue the anticipated loss.[126] It is true that at times the line between primary and secondary duties may be difficult to draw; and that in the Penalty Rule context the law may need to take a substantive approach to that distinction to avoid stultification. However, it does not follow that the distinction lacks normative substance. It is clearly normatively significant, for it divides lawful conduct from unlawful conduct.[127]

For another, the writer contends that some primary contractual obligations have the potential for 'overcompensation'—such as *all* duties to pay arising upon an event other than

---

[124] Text to nn 12-14 in ch 1.

[125] Allan Beever, *Rediscovering the Law of Negligence* (Hart Publishing 2007) 46 (emphasis added).

[126] Mindy Chen-Wishart, 'Controlling the Power to Agree Damages' in Peter Birks (ed), *Wrongs and Remedies in the Twenty-First Century* (Clarendon Press 1996) 277.

[127] Lionel Smith, 'Relief Against Forfeiture: a Restatement' (2001) 60 CLJ 178, 188 fn 28.

breach, or relating to a discount for prompt payment—and that if a clause 'involves objectionable over-compensation, it should be set aside in whichever form it is cast'.[128] Thus, Chen-Wishart seemingly suggests that a clause may be 'overcompensatory' even if it is not activated by an actual or substantive breach of duty. This reasoning is incoherent. In the private law, compensation is given to repair the loss that a wrong has caused. A wrong is an infringement of a right held by another. If no breach of duty exists, no wrong exists. Without a wrong, what precisely is one 'overcompensating'? The idea of compensation without a wrong is incoherent.

(2) The new

There is, however, another principled argument against the preventing 'overcompensation' rationale. In short, the traditional justification for the Penalty Rule is too narrow, for it excludes the possibility of a liquidated damages clause that relates to a legitimate, non-compensatory response. Properly understanding this point aids comprehension of the rule's true basis, which is outlined and evaluated in the next two chapters.

According to the preventing 'overcompensation' theory, Anglo-Australian courts solely apply the 'compensatory principle' when awarding general damages for breach of contract. This statement might once have held good. But times have changed. The private law now permits several other responses to this particular wrong, including disgorgement and substitution.

As to the former, in *A-G v Blake*, the House of Lords held that general disgorgement damages are exceptionally available for a breach of contract.[129] Conversely, the Australian position is presently unclear. In *Hospitality Group Pty Ltd v Australian Rugby Union Ltd*, a majority of the Full Court of the Australian Federal Court refused to recognise an award of disgorgement damages for breach of contract.[130] However, in the later of case *Dalecoast Pty Ltd v Guardian International Pty Ltd*, a majority of the West Australian Court of Appeal took a different view: these judges proceeded on the basis that the decision in *Blake* was correct, and that it applied in Australia.[131]

---

128 Mindy Chen-Wishart, 'Controlling the Power to Agree Damages' in Peter Birks (ed), *Wrongs and Remedies in the Twenty-First Century* (Clarendon Press 1996) 274.

129 [2001] 1 AC 268 (HL). See also *Esso Petroleum Co Ltd v Niad* [2001] EWHC 6 (Ch), [2001] All ER (D) 324 (Nov).

130 [2001] FCA 1040, (2001) 110 FCR 157 [157]-[159].

131 [2003] WASCA 142 [102]-[107].

Importantly, in *Hospitality Group*, Hill and Finkelstein JJ reasoned that ordering disgorgement damages would be inconsistent with the High Court's current approach to awarding general damages for breach of contract, which holds that such damages can only be compensatory.[132] However, general damages can clearly perform a function other than compensation. For example, nominal damages are unquestionably available for a breach of contract.[133] The response which nominal damages implement is controversial. However, whatever that response is, it is not compensation. Accordingly, the premise of their Honours' argument is undermined. The perceived monopoly of compensation is false.

As to the latter, rightly understood, English and Australian law have acknowledged the substitution response.

In England, several courts have awarded damages for breach of contract that are best interpreted as substituting for the value of the infringed performance right. Judges have not adopted the label 'substitutive damages'. However, the substitution response best explains a range of non-compensatory general damages awards made in breach of contract cases, including those involving contracts of sale; contracts for building work; and covenants not to build.[134] Leading English contract law works have acknowledged the substitutive damages thesis.[135]

One example will illustrate the point. In *Wrotham Park Estate Co Ltd v Parkside Homes Ltd*,[136] a developer acquired the proprietary right to land subject to a restrictive covenant imposed for the benefit of a neighbouring estate. The restrictive covenant prohibited development unless the current estate owner approved. In breach of the covenant, the developer built 14 houses and a road on the land. It then sold its proprietary right to the land, making a profit of £50,000. The estate owner sought an injunction restraining the building work, and mandating that the houses be demolished. Brightman J refused the injunction. However, the judge awarded general damages of £2,500. This sum represented the amount

---

132 *Hospitality Group Pty Ltd v Australian Rugby Union Ltd* [2001] FCA 1040, (2001) 110 FCR 157 [159].

133 eg *Luna Park (NSW) Ltd v Tramways Advertising Pty Ltd* (1938) 61 CLR 286 (HCA); *Surrey CC v Bredero Homes Ltd* [1993] 1 WLR 1361 (CA).

134 Robert Stevens, 'Damages and the Right to Performance: a Golden Victory or Not?' in Jason W Neyers, Richard Bronaugh and Stephen GA Pitel (eds), *Exploring Contract Law* (Hart Publishing 2009) 176-84, 186-94.

135 Edwin Peel, *The Law of Contract* (13th edn, Sweet and Maxwell 2011) [20-005].

136 [1974] 1 WLR 798 (Ch).

that the owner might reasonably have been demanded from the developer 'as a quid pro quo for relaxing the covenant'.[137]

The general damages award was not compensatory. The owner had suffered no loss. The developer's buildings did not diminish the value of the owner's proprietary right. Further, the owner did not lose any opportunity to bargain for the release of the covenant, because the judge expressly held that the owner would not have granted the developer any such release.[138] Any alleged loss, therefore, is fictional.[139] Nor was the award gain-based. Brightman J did not quantify the damages according to the developer's actual profit made as a result of the wrong, which was significantly higher than the amount awarded. A gain-based analysis, then, is equally as strained.

The most powerful analysis of the case is this.[140] Brightman J awarded the damages in lieu of specific relief, and so to substitute for the relevant right. Given a ready market for quantifying the value of this right did not exist, the judge constructed a reasonable, hypothetical bargain. So, Brightman J created a hypothetical bargain between a reasonable, willing seller and a reasonable, willing buyer for the seller to release the relevant right.[141] The judge then adopted a figure equal to a proportion of the developer's anticipated gain as a proxy for the value of the relevant right (and, thus, the right to a release). That is, Brightman J calculated this value at five per cent of the £50,000 profit that the developer expected to make (and did in fact make), being £2,500.

In Australia, courts have yet to consider expressly whether substitutive damages are available for a breach of contract. That said, many decisions are best understood as involving substitutive relief. A recent example is the Australian High Court's decision in *Clark v Macourt*.[142] Clark and Macourt were doctors specialising in providing assisted reproductive technology services. Clark entered into a contract with Macourt to buy certain assets of his practice, including a stock of frozen donated sperm 'straws'. The contract price was approximately $390,000. Macourt warranted that the sperm donors had been tested and identified in accordance with specified guidelines. Macourt delivered 3,513 sperm straws.

---

137 *Wrotham Park Estate Co Ltd v Parkside Homes Ltd* [1974] 1 WLR 798 (Ch) 815.

138 *Wrotham Park Estate Co Ltd v Parkside Homes Ltd* [1974] 1 WLR 798 (Ch) 815.

139 *Surrey CC v Bredero Homes Ltd* [1993] 1 WLR 1361 (CA) 1369.

140 Robert Stevens, 'Damages and the Right to Performance: a Golden Victory or Not?' in Jason W Neyers, Richard Bronaugh and Stephen GA Pitel (eds), *Exploring Contract Law* (Hart Publishing 2009) 192.

141 *Wrotham Park Estate Co Ltd v Parkside Homes Ltd* [1974] 1 WLR 798 (Ch) 815-16.

142 [2013] HCA 56, (2013) 253 CLR 1.

However, 1,996 straws did not comply with the warranty given, and were unusable. Clark purchased 1,996 replacement, contractually-compliant sperm straws from another supplier, at a cost of about $1 million. Further, Clark charged each patient a fee that covered the costs she had incurred in acquiring the replacement straws used to treat that patient. In other words, she recouped her costs, but did not make a profit. Macourt sued to recover the outstanding balance of the contract price; Clark counterclaimed for damages for breach of contract.

Before the High Court, the only issue was the proper basis for quantifying Clark's general damages. Clark argued that Macourt's breach of contract meant that the value of the sperm straws as assets was less than it would have been if he had kept his promise. Further, she contended that damages comprising that difference in value should be assessed at the breach date. In reply, Macourt argued that Clark's claim was in fact to recover the costs associated with the procurement of replacement sperm. Also, he submitted that any damages quantifying that loss should be assessed at the trial date.

The majority of the High Court held that Clark was entitled to general damages based on the amount that she would have had to pay the alternative supplier, as at the date of completion, to buy 1,996 replacement sperm straws. Put differently, Clark was awarded the value of 1,996 warranty-compliant straws as at the breach date.

Consider a loss-based analysis. Assume that, contrary to the majority's view, the damages claim as framed was properly for consequential loss. If so, Clark's initial factual detriment was her cost of purchasing donor sperm straws to replace the non-compliant sperm straws. However, as stated, before the trial date Clark recouped from her patients the expenditure she had incurred in obtaining the replacement straws. As Gageler J pointed out when dissenting, factually she was left no worse off at the end of the story.[143] As a matter of doctrine, Clark's recoupment constituted a direct compensating advantage that 'wiped out' the initial factual detriment that she had suffered.[144] To explain, the compensating advantage was 'direct' because the recoupment derived from the acts Clark took precisely *to avoid* the detrimental factual consequences of Macourt's breach.[145] In other words, the mitigating act comprised Clark acquiring the replacement stock and using it in her practice, and the recovered costs arose 'out of [that] act of mitigation itself'.[146] As the advantage was direct, its

---

[143] *Clark v Macourt* [2013] HCA 56, (2013) 253 CLR 1 [71].

[144] *Erie County Natural Gas and Fuel Co Ltd v Carroll* [1911] AC 105 (PC) 119.

[145] Andrew Burrows, *Remedies for Torts and Breach of Contract* (3rd edn, OUP 2004) 158.

[146] Harvey McGregor, *McGregor on Damages* (19th edn, Sweet and Maxwell 2014) [9-115]. See *British Westinghouse Electric and Manufacturing Co Ltd v Underground Electric Railways Co of London Ltd* [1912] AC 673 (HL) 689-90.

value had to be deducted from the initial factual detriment. Accordingly, Clark did not in fact suffer any net loss. It follows that the substantial general damages award was not compensatory.

Contrast a rights-based approach. On this view, the damages award sought to vindicate Clarke's infringed performance right by substituting for the value of that right. If so, the Court's reasoning, and the outcome, is explicable. Thus, the majority held that the damages claim as framed was one for the difference in market value between the straws promised and those delivered, which must be assessed at the breach date.[147] Put another way, it was a claim based upon the value of the infringed performance right. Further, the majority held that the cost of purchasing replacement straws was not *directly* relevant to Clark's claim. Rather, the claim was for the value of the straws that should have been delivered, and the amount paid to the alternate supplier was merely evidence of that value.[148] As such, the incurred cost was only relevant because it provided an 'appropriate proxy'[149] for the value of the infringed right. In addition, Keane J suggested that a general damages claim of the instant type displays the following features: that matters arising after the breach date have 'no bearing' on the claimant's right and defendant's duty; that the compensating advantages rule is 'immaterial'; and that the mitigation doctrine is irrelevant.[150] Of course, substitutive damages are assessed at the breach date, and are not subject to the general law principles limiting compensatory damages.

In short, if the case concerned compensation, it was wrongly decided; if it concerned substitution, it was rightly decided. The latter view, then, is better.

Therefore, as best interpreted, in Anglo-Australian private law compensation is not the only response available to a breach of contract. The implications for the preventing 'overcompensation' theory are devastating. As illustrated, the theory is predicated on three central premises: that in awarding general damages for breach of contract, Anglo-Australian courts always apply the compensatory principle; that the agreed damages regime is an analogue of the general damages regime; and that in agreeing damages, the parties must apply the compensatory principle, and must imitate a judicial compensatory remedy. However, if the hegemony of compensation has been broken at general law, then only the second premise

---

[147] *Clark v Macourt* [2013] HCA 56, (2013) 253 CLR 1 [13], [25], [103], [109], [111].

[148] *Clark v Macourt* [2013] HCA 56, (2013) 253 CLR 1 [51]-[52], [100] (Counsel's submissions).

[149] *Clark v Macourt* [2013] HCA 56, (2013) 253 CLR 1 [39]. See similarly [20]-[21], [138].

[150] *Clark v Macourt* [2013] HCA 56, (2013) 253 CLR 1 [110], [128]-[129], [133]-[134], [143]. See also [37] (Crennan and Bell JJ).

remains true. If judges are entitled to award substitutive, compensatory, restitutionary and disgorgement damages for breach of contract, and the parties' agreed damages must follow in the footsteps of a general damages remedy, then the compensatory principle cannot rationally restrict the parties' remedial choice. Accordingly, the availability of other responses at general law proves that preventing 'over*compensation*' cannot provide a satisfactory rationale for the rule.

IV Conclusion

The preventing 'overcompensation' basis derives from the modern law's unfortunate obsession with loss. In fact, other responses—including substitution and disgorgement—are available for breach of contract. If one takes an associated view of agreed damages, then a desire to prevent 'over*compensation*' cannot justify the Penalty Rule. It is too narrow. Properly understood, the theory is one of preventing 'overcompensation', 'oversubstitution', and so on.

It is important to note that, as used in this context, the term 'overcompensation' is an abuse of language. The noun 'overcompensation' is derived from the verb 'overcompensate'. The Oxford English Dictionary defines the latter as follows: 'Take excessive measures in attempting to correct or make amends for an error, weakness, or problem.'[151] This definition does not accord with the meaning of the word as used in this setting. Also, the word 'overcompensation' as here adopted denotes an amount that represents *more than* compensation. If so, the term is unintelligible: an amount that goes beyond compensation sensibly cannot be labelled 'compensation' at all.

4 Conclusion

The literature contains three leading theories purportedly explaining why courts invalidate penalty clauses. This chapter has revealed that each putative justification is problematic, as follows.

First, the substantive unconscionability rationale fails the coherence criterion for assessing private law theories. It also displays intractable problems of fit, transparency and principle; and raises practical concerns. Second, the efficiency theory fails the transparency criterion for assessing private law theories. Issues of coherence, principle and practice also plague the rationale. Presently, no cure exists. Finally, the classic preventing

[151] Oxford Dictionaries <www.oxforddictionaries.com/definition/english/overcompensate> accessed 12 January 2015.

'overcompensation' rationale is in fact quite robust. However, it presents one problem of principle, and a related issue of terminology, that is fatal.

It follows that there is no *obvious* justification for the Penalty Rule. This has led some Anglo-Australian courts to give up the search for a theoretical rationalisation. For instance, in *Robophone Facilities Ltd v Blank*, Diplock LJ declared: 'I make no attempt, where so many others have failed, to rationalise this … rule. It seems to be *sui generis*.'[152] Similarly, in *Murray v Leisureplay plc*, Arden LJ held that the reasons for the rule 'may be pragmatic rather principled'.[153] Recently, in *Kellas-Sharpe v PSAL Ltd*, Fryberg J held that he agreed 'with those who think the rule anomalous'.[154]

These comments are extraordinary. In describing an instance of legal unenforceability as *sui generis*, 'pragmatic rather than principled', and anomalous, these judges intend that no general principle can justify this instance. That is to admit that the private law is *ad hoc*, and therefore comes very close to admitting that this rule is unjustified. Indeed, leading academics have seemingly reached precisely that view. For example, Heydon and Leeming have suggested that the doctrine is merely a collection of 'arbitrary and pointless dogmas'.[155] Further, Peel has argued that 'there is now no basis at all for the [Penalty Rule]. Increasingly, it looks like an anomaly.'[156] Likewise, Worthington has concluded that it 'seems impossible to rationalize' the rule.[157] In short, the traffic is all one way.[158] Based on the theories this chapter has considered, these statements may indeed turn out to be true. But it is a deeply unattractive view that we should be reluctant to accept without examining all the alternatives.

What is required, then, is an alternative theory. The next chapter provides precisely that. It argues that, viewed from a taxonomic perspective, several types of agreed damages clauses exist, including agreed compensatory, substitutive and disgorgement damages clauses. Each type of clause necessarily may be penal in some instances. A satisfactory justification must focus on the particular feature that renders each type of clause penal. In each case, a clause will be invalidated only if it is *excessive* in light of its particular function. The following

---

[152] [1966] 1 WLR 1428 (CA) 1446.

[153] [2005] EWCA Civ 963, [2005] IRLR 946 [29].

[154] [2012] QCA 371, [2013] 2 Qd R 233 [57].

[155] JD Heydon and MJ Leeming, *Cases and Materials on Equity and Trusts* (8th edn, LexisNexis Butterworths 2011) [17.1].

[156] Edwin Peel, 'The Rule Against Penalties' (2013) 129 LQR 152, 156.

[157] Sarah Worthington, *Equity* (2nd edn, OUP 2006) 227.

[158] See also Janet O'Sullivan, 'Lost On Penalties' (2014) 73 CLJ 480, 483.

chapter argues that this additional, excessive element is punitive. Hence, in truth, the Penalty Rule's rationale is one of preventing consensual punishment, an analogous rationale to that which underpins the general rule against punitive damages. Penalty clauses, then, are a punitive damages analogue. Further, the next chapter demonstrates that corrective justice norms justify the preventing consensual punishment principle.

This argument is not novel. Burrows once argued that the Penalty Rule was justified on the following terms:

> compensation rather than punishment is such a central feature of damages … that it would be wrong to allow parties to oust that regime. Punishment belongs to the criminal and not the civil law. And so while the courts should and do allow the parties to make their own assessment of compensation, it is unjustifiable … for the [claimant] to insist on punishing the defendant.[159]

Although the private law has since revealed the false monopoly of compensation, the main insight is crucial. Rediscovering it forms the key to understanding, and justifying, the rule.

[159] Andrew Burrows, *Remedies for Torts and Breach of Contract* (2nd edn, Butterworths 1994) 331 (footnote omitted).

8 EXPLANATION

Holmes once famously said that 'The life of the law has not been logic: it has been experience'.[1] If that is right, then a rule's longevity alone provides sufficient reason for its continued existence. There is no need to justify it. Another view exists, however. Birks thought the private law needed 'more logic and less experience'.[2] The writer always asked how a rule is best classified, and then whether it can be justified. If that is right, then classification is critical—for in many cases one must classify a rule before one can justify it—and justification is essential.

Birks's view is preferable. Accordingly, this chapter seeks to classify, and then justify.

This chapter provides a two-level theory of the Penalty Rule. It is divided into two sections. In the first section, the chapter outlines a *classificatory* theory. This constitutes the low level of the overall theory, and it reveals the grounded principle in play in this area. Thus, this section recapitulates on the extended Birksian private law map, and then situates liquidated damages clauses and penalty clauses within that schema. That exercise reveals the rule's true underlying principle. In the second section, the chapter presents a *corrective justice* theory. This comprises the high level of the overall theory, and it enunciates a more profound theoretical explanation of the law. Hence, this section illustrates that the rule's underlying principle instantiates a higher normative rationale, which provides the rule's ultimate theoretical justification.

This chapter demonstrates that the Penalty Rule encapsulates a principle of preventing consensual civil punishment, and corrective justice norms justify that principle. Accordingly, a classificatory, corrective justice theory provides an understanding that explains the Penalty Rule rationally, setting out good or legitimate reasons why the rule is as it is.[3] In other words, this chapter establishes that the Penalty Rule *instantiates* corrective justice. It follows that the rule is not only readily defensible, but it has a powerful logic and strength.

1 Classificatory theory

---

1 Oliver Wendell Holmes, *The Common Law* (Little Brown 1881) 1.

2 Peter Birks, 'Equity in the Modern Law: An Exercise in Taxonomy' (1996) 26 UWAL Rev 1, 5.

3 John Gardner, 'What is Tort Law For? Part 1: the Place of Corrective Justice' (2011) 30 Law & Phil 1, 3.

A The private law map

As Chapter 1 discussed, classification concerns mapmaking. It involves working out what are the various parts of the private law, and then making clear how they fit together. Classification is important because it promotes understanding. As stated, classification is also critical because it provides a preliminary step to justification.[4] So, in order properly to justify the Penalty Rule, it is necessary first to comprehend its place on the private law map.

The Birksian private law map posits that four causative events potentially generate one of five responses. However, Chapter 1 demonstrated that the Birksian taxonomy is incomplete. A complete, extended classificatory scheme was represented in a table, namely Table A.[5] The extended Birksian private law map posits that four causative events (consent, wrongs, unjust enrichment and 'other events') potentially elicit one of seven responses (substitution, compensation, restitution, disgorgement, punishment, perfection and 'other functions').

The following questions arise: what is the causative event that generates a right to liquidated damages? What is the function of that right when it arises? Similarly, if the Penalty Rule did not exist and a penalty clause was enforceable, what would be the causative event that would elicit a right to a penal sum? What would be the right's function? In summary, what are the map co-ordinates for a right to liquidated damages, and for a right to a penalty? These questions are examined in order.

B The coordinates for liquidated damages clauses

This section examines the taxonomy of liquidated damages: that is, the possible causative events that lead A to obtain against B a right to liquidated damages, and the potential functions of that right when it arises. The legal materials do not even suggest that liquidated damages arise in response to unjust enrichment and 'other events'. Accordingly, this section considers only the two *plausible* causative events: wrongs and consent. As to responses, no one has ever contended that a right to liquidated damages displays a punitive function, or a miscellaneous 'other function'. Accordingly, this section passes over those categories.

I The causative event

---

[4] Alternatively, it might be contended that classification is an *integral part* of justification: in particular, that classification provides the first step in any justificatory process. Nothing substantive turns on this difference.

[5] Table following n 57 in ch 1.

(1) Wrongs

The orthodox view is that a right to liquidated damages arises from a wrong. Chapter 1 noted that, on a rights-based approach, where a defendant violates a claimant's primary claim right, a secondary claim right arises. So, where Y breaches his contract with X, X obtains a secondary right against Y for general damages. Similarly, it is thought that where the contract contains a valid agreed damages clause, upon breach A obtains a secondary right against B for liquidated damages.[6] A's secondary right, sometimes called a 'remedial' right, thus arises from the general law wrong of B's breach of contract.

Much support for this view exists. For instance, Chen-Wishart argues that agreed damages clauses involve 'secondary (remedial)' obligations, and that such clauses provide 'agreed *remedies* for breach of the primary obligations'.[7] The writer refers to *Photo Production Ltd v Securicor Transport Ltd*, where Lord Diplock held that primary rights are born of the contract, whereas secondary rights are born of the wrong of breach of contract; and that 'secondary obligations … arise by implication of [general] law'.[8] Some have interpreted his Lordship's judgment as holding that liquidated damages clauses 'confer secondary rights', and hence as supporting the view that the right to liquidated damages arises from a wrong.[9] On this view, then, a liquidated damages clause simply modifies an obligation imposed by law. Thus, Carter contends that the duty to pay liquidated damages arises from the wrong of breach of contract; and that one is dealing with the usual general law secondary obligation to pay damages, though that obligation is 'quantified' in nature.[10]

For at least three reasons, the relevant causative event cannot be a wrong.

To start, in *Photo Production*, Lord Diplock in fact held that the 'remedial' right to liquidated damages is born *of the contract*, not of the wrong. His Lordship stated that although secondary obligations normally arise by implication at general law, 'The contract, however, is just as much the source of secondary obligations as it is of primary obligations'.[11] This statement, however, is potentially misleading. As illustrated below, although the right to

---

[6] Recall that this proposition formed one of the key premises of the preventing 'overcompensation' thesis.

[7] Mindy Chen-Wishart, *Contract Law* (4th edn, OUP 2012) [14], [14.3.2.1] (original emphasis).

[8] [1980] AC 827 (HL) 848-49.

[9] Rafal Zakrzewski, *Remedies Reclassified* (OUP 2005) 111.

[10] *Carter on Contract* [43-100] (service 37, April 2014).

[11] *Photo Production Ltd v Securicor Transport Ltd* [1980] AC 827 (HL) 849. See also *GWC Property Group Pty Ltd v Higginson* [2014] QSC 264 [19].

liquidated damages adopts the *normative structure* of a general law secondary right, it is nonetheless contractual—hence consensual—in origin, and hence a *primary* right. That said, the presently important point is that, in Lord Diplock's view, the right to liquidated damages does not arise from the general law wrong of breach of contract.

Next, the appropriate response to the general law wrong of breach of contract is assessed *ex post*, at the breach date or later. Conversely, the causative event that elicits the right to liquidated damages demands that the response is assessed *ex ante*, at the contract date.

Lastly, for a claimant to recover a substantial sum for a wrong, she must prove matters beyond a mere breach of duty (for example, she must prove the value of her infringed right, or her consequential loss). By contrast, if A proves that a liquidated damages clause exists and that B has breached it, she may recover without further ado.

In brief, a general law wrong does not generate a right to liquidated damages.

(2) Consent

As Lord Diplock suggested in *Photo Production*, it is the parties' agreement that creates a claim right to liquidated damages. Put differently, the right to liquidated damages arises from consent. At least three reasons might be given.

First, in *Rothenberger Australia Pty Ltd v Poulsen*, Barrett J held that liquidated damages are 'recoverable in satisfaction of a right of recovery created by the contract itself and accruing by reason of breach'.[12] A recovery right 'created by the contract' is generated by consent. Similarly, as Burrows has noted, in cases involving agreed damages clauses the true concern is with 'the validity of [B's] promise'.[13] As this promise is given consensually, one is dealing with the causative event of consent.

Second, consent is essential to enforce a liquidated damages clause, and if it does not exist the fact that a breach of contract has occurred is irrelevant. So, suppose that A exerts illegitimate pressure upon B to enter into a contract with her. The contract contains a liquidated damages clause. B then breaches the contract, triggering the liquidated damages clause. If B avoids the agreement for a lack of consent (due to duress), A cannot recover from B any liquidated damages, despite the fact that a breach occurred at a time when the contract was valid.

---

[12] [2003] NSWSC 788, (2003) 58 NSWLR 288 [27].

[13] Andrew Burrows, *Remedies for Torts and Breach of Contract* (3rd edn, OUP 2004) 441.

Finally, taxonomically speaking, the most significant feature of a debt is that it most commonly[14] arises from consent. As Chapter 6 showed, liquidated damages are generally in the nature of a debt, which is the product of consent;[15] and not in the nature of damages, which are the product of a general law wrong.

One fundamental point must be highlighted. Zakrzewski argues that a duty to pay under a liquidated damages clause is 'analytically a primary obligation' that is 'merely triggered by … the breach of another contractual obligation'.[16] Accordingly, even if the event triggering B's duty under a liquidated damages clause is a breach of contract, B's duty is nonetheless a primary duty, because it arises from the parties' consent. However, B's primary duty exhibits the *same structure or anatomy* as a secondary duty that arises from a general law wrong. Taking a formalist approach,[17] it is the 'structural' private law concepts that determine the result of particular types of cases, and if two such 'structural' concepts bear a like form, the private law must treat those concepts similarly. It follows that the private law will treat B's primary duty in the same way as a defendant's general law secondary duty, because B's primary duty exhibits the same genetic make-up as a general law secondary duty.

Thus, B's primary duty takes the same shape as a general law secondary duty that arises upon the wrong of breach of contract. However, this does *not* mean that a right to liquidated damages arises from a general law wrong. Rather, the situation is one where a consent-based event 'picks up' on the structure of a wrong: the waters of consent are poured into a wrongs-shaped vase. In other words, this particular manifestation of consent bears the anatomy of a general law wrong. In the case of liquidated damages, then, everything occurs in the land of consent. This point is crucial.

II The response

(1) Compensation

Carter and Courtney define a liquidated damages clause as 'an agreement between the parties

---

[14] A debt may arise due to an unjust enrichment. For example, a claim right to recover a mistaken payment is an action in debt.

[15] Only a liquidated damages clause involving a duty to pay money may be enforced as a debt. That said, most liquidated damages clause takes this form. As a result, *in most cases*, an action to recover liquidated damages is in the nature of a debt claim.

[16] Rafal Zakrzewski, *Remedies Reclassified* (OUP 2005) 110 (footnote omitted).

[17] Text following n 37 in ch 1.

as to the compensation payable on breach …'.[18] This is the 'core case'. Most commonly, the function of A's right to liquidated damages is to repair an anticipated loss. In such a case, correctly labelled the right in play is one to liquidated compensatory damages: that is, 'contracted compensatory sums that are not penalties'.[19]

In the agreed damages context, courts have recognised the compensatory response for hundreds of years. The majority of cases to which this thesis refers could be cited in support of the proposition. Nothing more need be said.

(2) Perfection

The perfection response seeks substantially to achieve the very object of a causative event. A court order is perfectionary, then, if it requires a party to do, or not to do, the very thing she undertook to do, or not to do.

It is possible to contend that the function of A's right to liquidated damages is to achieve the object of the parties' consensual agreement. The reasoning might run thus. A's primary right to liquidated damages is the concomitant of B's primary duty to do the very thing undertaken, such as to pay a liquidated sum. It follows that this right, and the court's order that replicates that pre-existing right, is perfectionary. Some scholars apparently take this view. For instance, Burrows seemingly suggests that a court order to pay liquidated damages is perfectionary. The writer considers that a liquidated damages award is 'the same type of judicial remedy as the award of an agreed price',[20] as both remedies compel performance of a positive contractual obligation.

This view has some purchase. If a liquidated damages clause is engaged, B's duty to pay, transfer or extinguish is a primary duty, such that a court order is perfectionary in that it enforces A's corresponding primary right. Accordingly, it is possible to classify *all* liquidated damages clauses as instantiating the perfection response. However, this classification is too blunt. It would conceal important distinctions between different kinds of rights to liquidated damages, inhibiting a full understanding. It is better to classify any right to liquidated damages by reference to the more granulated response that the right embodies, for instance substitution, compensation or disgorgement. Such categorisation is more transparent, as it

---

[18] JW Carter and Wayne Courtney, 'Indemnities against Breach of Contract as Agreed Damages Clauses' [2012] JBL 555, 558.

[19] *Rothenberger Australia Pty Ltd v Poulsen* [2003] NSWSC 788, (2003) 58 NSWLR 288 [18].

[20] Andrew Burrows, *Remedies for Torts and Breach of Contract* (3rd edn, OUP 2004) 440.

more accurately describes the nature of the right that arises. For example, assume B is obligated under a liquidated damages clause to pay money to A to repair an anticipated harm. A perfectionary classification is insufficiently fine-grained to explain satisfactorily what the private law is doing. Conversely, a compensatory classification provides a more faithful explanation of how A's right functions.

Thus, a right to liquidated compensatory damages might be labelled as perfectionary *and* compensatory. However, to prevent an overlap, this right must only bear one set of coordinates. As it is best viewed as a right to consensual compensation, it must live in the box in Table A where consent and compensation intersect.

(3) Substitution

Substitution seeks to vindicate an infringed right by substituting for the value of that right. In several seminal cases, courts have recognised that the function of A's right to liquidated damages may be to substitute for value of the infringed performance right. For instance, in *Comr of Public Works v Hills*, Lord Dunedin held that an agreed damages amount is valid if it represents 'a genuine pre-estimate of the creditor's probable or possible interest in the due performance of the principal obligation'.[21] In other words, a right to liquidated damages arises if the clause appropriately pre-assesses the value of A's infringed performance right.

To date, liquidated substitutive damages clauses have mostly arisen in contracts involving public authorities. However, contracts between private parties have also contained such clauses. These scenarios are discussed in turn.

Beginning with contracts concerning public authorities, public entities are not usually involved in commercial, profit-generating enterprises. Accordingly, if a public authority enters a contract for the construction of public works (for instance, of a dam or a road), or the supply of public goods (for instance, of warships), generally it will not suffer a financial loss from non-completion or delayed completion. Consequently, it is said to be normally 'impossible to quantify the loss'[22] which non-performance or delay will cause. As Chapter 5 demonstrated, in such cases the anticipated 'loss' is impossible to estimate because it does not in fact exist.

Anglo-Australian case law support the view that, in contracts involving public entities, a right to liquidated damages may be substitutive in nature.

---

[21] [1906] AC 368 (PC) 375-76. See similarly *Lamson Store Service Co Ltd v Russell Wilkins and Sons Ltd* (1906) 4 CLR 672 (HCA) 681-82, 686, 688-89.

[22] Stephen A Smith, *Atiyah's Introduction to the Law of Contract* (6th edn, Clarendon Press 2005) 392.

The best English example is the House of Lords's decision in *Clydebank Engineering and Shipbuilding Co Ltd v Don Jose Ramos Yzquierdo y Castaneda*.[23] As Chapter 5 examined that judgment in detail,[24] the facts and decision of the case will not be rehearsed here. It suffices to say that, as best interpreted, the House of Lords upheld the agreed substitutive damages clause in the contract between the Spanish Government and the shipbuilder because the agreed damages amount was not disproportionate to the anticipated value of the infringed performance right. In that case, the value of the right was quantified adopting a reasonable, hypothetical bargain approach.

An illuminating Australian decision is *State of Tasmania v Leighton Contractors Pty Ltd*.[25] Leighton agreed to design and construct a highway for the Tasmanian State Government. The contract included an agreed damages clause stipulating that if Leighton was late in completing the works, it would be obliged to pay 'liquidated damages' at the rate of $8,000 per day for the period of its delay. The Commonwealth Government was contractually obligated to reimburse the State Government for the project costs. Leighton completed 229 days late. The State Government purported to extinguish Leighton's right to be paid for work it had completed in the sum of $1.83 million. In response, Leighton alleged that the clause was penal.

The Full Court of the Tasmanian Supreme Court held that the clause was valid. The Court held that, in arriving at the agreed damages amount, the Government was entitled to take account of two 'separate matters': any anticipated 'loss of public utility or delay in access to infrastructure'; and any expected 'direct costs' of supervising an overrun contract (including services fees and administrative costs).[26] Properly analysed, both components are substitutive in nature.

As to the 'loss' of public utility, it is artificial to classify items such as 'loss of public amenity',[27] the public's inability to use the road, and delays to the public in accessing infrastructure as losses that *the Government* would incur. As regards these items, one cannot correctly say that the Government would be factually worse off as a result of Leighton's breach. More accurately, upon Leighton's default, the Government's performance right would be infringed and *the community at large* would suffer these (non-pecuniary) losses, which are consequent upon the infringement of that right. Indeed, the Full Court acknowledged this,

---

[23] [1905] AC 6 (HL).

[24] Text accompanying and following n 80 in ch 5.

[25] [2005] TASSC 133, (2005) 15 Tas R 243.

[26] *State of Tasmania v Leighton Contractors Pty Ltd* [2005] TASSC 133, (2005) 15 Tas R 243 [13], [31].

[27] *State of Tasmania v Leighton Contractors Pty Ltd* [2005] TASSC 133, (2005) 15 Tas R 243 [31(1)].

stating that a breach of a public works contract 'might result in loss or harm to public convenience'.[28]

As to the 'loss' of the direct costs, as stated the Commonwealth Government was contractually obligated to reimburse the State Government for the project costs. Consequently, the trial judge had held that the State Government would not suffer any anticipated loss.[29] Put differently, in any event factually the State Government would be left no worse off. It followed that its anticipated legally-recoverable loss was nothing, such that the agreed damages amount was clearly excessive, and the clause penal. Compare the Full Court's view. It held that the fact of reimbursement did not matter: although the Commonwealth Government would ultimately suffer the factual detriment, it was the State Government's performance right that would be violated, and the agreed damages clause was valid because it properly reflected the anticipated value of that infringed right.

Turning to contracts between private parties, it has been held that a right to liquidated damages may legitimately function by vindicating an anticipated infringed right. A well-known example is Lord Goff's judgment in *Alfred McAlpine Construction Ltd v Panatown Ltd*.[30] Panatown employed McAlpine to design and construct an office building on land owned by UIPL, a related entity of Panatown. The construction contract included an agreed damages clause providing that if McAlpine was late in erecting the building, it would be obliged to pay agreed damages at the rate of £35,000 per week for the period of its delay. Panatown alleged that the building was seriously defective, and completion was delayed. Given UIPL held the proprietary right to the land, it was UIPL that had suffered any factual loss consequent upon McAlpine's breach of contract. Despite this, Panatown claimed general damages for defective work, general damages for delay, and agreed damages for delay.

By a bare majority, the House of Lords held that Panatown could recover only nominal damages. Lord Goff, who dissented, was the only Law Lord to consider Panatown's agreed damages claim. His Lordship noted the trial judge's decision that the agreed damages clause was penal.[31] Taking a strict loss-based view, the trial judge had held that the agreed damages amount was clearly extravagant when compared with the factual loss that Panatown was expected to suffer from the delay, which was nothing. However, Lord Goff adopted a rights-

---

[28] *State of Tasmania v Leighton Contractors Pty Ltd* [2005] TASSC 133, (2005) 15 Tas R 243 [38].

[29] *State of Tasmania v Leighton Contractors Pty Ltd (No 3)* [2004] TASSC 132 [241].

[30] [2001] 1 AC 518 (HL).

[31] *Alfred McAlpine Construction Ltd v Panatown Ltd* (1998) 58 Con LR 46 (QB) 80.

based approach, and viewed the clause as an agreed substitutive damages clause. His Lordship held that an employer has a 'right to the performance by the contractor of his obligation as to time' as well as to performance in accordance with the contractual specification, such that the clause was valid on the ground that the agreed damages sum represented 'a genuine pre-estimate of the damage suffered by Panatown by reason of the delay in receiving the benefit of the building work to be performed by McAlpine under the building contract'.[32] Regarding quantification, Lord Goff held that the 'appropriate yardstick' for measuring the anticipated value of the employer's infringed right to timely performance was 'the anticipated profitability of the development'.[33]

Thus, substitution is an accepted response in the liquidated damages context. A may claim from B liquidated damages providing a substitute for the anticipated value of the infringed performance right.

(4) Disgorgement

The disgorgement response involves divesting a gain. Generally speaking, it means a party must 'give up' profits he has wrongfully made. Can the function of A's right to liquidated damages be to require B to disgorge an anticipated factual gain?

An example: soldier Y agrees with government X that Y will not disclose any sensitive information relating to Y's official duties after leaving the armed forces. The contract contains an agreed damages clause stating that if Y breaches his non-disclosure obligation, he must pay to X the sum of £100,000. The parties quantify the agreed damages amount according to the profit that Y is likely to make from the breach. X will not suffer any loss as a result of the breach. Is the clause valid? On a loss-based view, the clause is penal. Conversely, the better view is that an appropriately drafted agreed disgorgement damages clause is valid. Various Commonwealth courts have supported this proposition. Two examples follow.[34]

For one, in Australia it has been held that, in principle, a right to liquidated damages may be profit stripping. The relevant case is *Harris v Digital Pulse Pty Ltd*.[35] Digital provided multi-media services to clients. Two Digital employees, Harris and Eden, worked in marketing and web design respectively. It was an express term of their employment contracts that each

---

[32] *Alfred McAlpine Construction Ltd v Panatown Ltd* [2001] 1 AC 518 (HL) 554-55.

[33] *Alfred McAlpine Construction Ltd v Panatown Ltd* [2001] 1 AC 518 (HL) 555.

[34] See also *Surrey CC v Bredero Homes Ltd* [1993] 1 WLR 1361 (CA) 1364.

[35] [2003] NSWCA 10, (2003) 56 NSWLR 298.

of them would not compete with Digital. Whilst still employed by Digital, the two employees secretly incorporated a separate company, Juice-D, to compete with Digital. The employees then diverted business opportunities from Digital to Juice-D. Digital sued its former employees for breach of contract and breach of fiduciary duty. The trial judge held that the employees had 'breached their contractual and fiduciary duties of loyalty',[36] and awarded Digital disgorgement damages and punitive damages. On appeal, the New South Wales Court of Appeal held that punitive damages are unavailable for a breach of fiduciary duty.

In his judgment, Spigelman CJ hypothesised an agreed damages clause stating that 'the employer could recover, for a relevant breach … an account of profits plus $10,000. It is clear that the … $10,000 would be struck down as a penalty.'[37] Mason P 'readily agreed' that the amount 'over and above' the disgorgement sum would be penal.[38] Unfortunately, the Chief Justice's example was lacking, in that any agreed damages clause permitting recovery of anticipated profits must *quantify* the relevant amount. Without the necessary quantification feature, the clause is not an agreed damages clause at all.[39] The dictum, however, is crucial. Spigelman CJ recognised that an agreed damages clause that appropriately estimates B's likely gains will be valid. The clause will only be invalid if it contains an excessive, 'over-disgorgement' component. Even then, it will be invalid only to the extent of that excess. Thus, in Australia, a liquidated disgorgement damages clause is a valid concept.

Heydon JA impliedly supported the Chief Justice's reasoning. His Honour defined a 'penalty' and 'punishment' as exacting a 'money sanction greater than that which is needed either to give full compensation for loss or *full disgorgement of gain*'.[40] This suggests that the 'penal' element of a financial sanction is limited to the component that goes beyond any profit-stripping sum.

For another, consider the New Zealand case of *R v A-G for England and Wales*.[41] R was a member of the SAS. He entered an agreement with the Crown. Clause 1 contained a confidentiality clause prohibiting R from disclosing any information relating to the SAS's activities. Clause 4 provided that R would assign to the Crown 'all rights accruing to [him] and

---

36 *Harris v Digital Pulse Pty Ltd* [2003] NSWCA 10, (2003) 56 NSWLR 298 [70]. See also [475].

37 *Harris v Digital Pulse Pty Ltd* [2003] NSWCA 10, (2003) 56 NSWLR 298 [58].

38 *Harris v Digital Pulse Pty Ltd* [2003] NSWCA 10, (2003) 56 NSWLR 298 [187].

39 Text to n 15 in ch 1.

40 *Harris v Digital Pulse Pty Ltd* [2003] NSWCA 10, (2003) 56 NSWLR 298 [407] (emphasis added).

41 [2002] 2 NZLR 91 (NZCA), affd [2003] UKPC 22, [2003] EMLR 24.

arising out of' a breach of the clause.[42] An explanatory memorandum reiterated that the Crown was contractually entitled to strip R of any profits that he made from a breach. After leaving the army, R sought to publish his memoirs, which revealed information concerning SAS operations. The New Zealand Court of Appeal refused an injunction restraining R from publishing his book, but ordered R to pay disgorgement damages. On appeal, the Privy Council upheld that decision.

Relevantly, in the Court of Appeal Tipping J held that clause 4 provided a contractual right to gain-based relief. His Honour further held that clause 4, which contained a 'contractually provided remedy of … assignment of, profits', was valid and enforceable.[43] Admittedly, it is unclear whether a clause referring to the assignment of 'all rights' suffices to satisfy the quantification requirement. Some writers seemingly take the view that it does.[44] If so, the decision suggests that courts will give effect to an agreed disgorgement damages clause in appropriate circumstances.

In short, a right to liquidated damages may operate by divesting a gain.

(5) Restitution

Restitution is a gain-based response that seeks to reverse a transfer of value between the parties. Generally speaking, it means a party must 'give back' a negative benefit that he has wrongfully obtained. Restitution is a less extreme gain-based response than disgorgement. As such, if Anglo-Australian private law is willing to enforce an agreed disgorgement damages clause, it necessarily follows that it will sanction an appropriately drafted agreed restitutionary damages clause. In principle, then, A should be entitled to recover from B liquidated restitutionary damages.

In conclusion, the function of A's right to liquidated damages legitimately may be any of the following: substitution, compensation, restitution or disgorgement.

C The coordinates for penalty clauses

This section examines how penalties are properly to be classified. As Chapter 6 established, a true penalty clause is partially unenforceable. Thus, where a contract contains a penalty clause, the clause creates a right to a penalty, but part of that right—the excessive component—is

---

42 R *v A-G for England and Wales* [2003] UKPC 22, [2003] EMLR 24 [9].

43 *A-G for England and Wales v* R [2002] 2 NZLR 91 (NZCA) [113].

44 eg John McGhee (gen ed), *Snell's Equity* (33rd edn, Sweet and Maxwell 2014) [13-013].

unenforceable. For expository purposes, it must be assumed that the Penalty Rule did not exist. Supposing that the right to a penalty is wholly enforceable and that the entire penal sum is recoverable, two questions arise. First, what is the causative event that elicits the right to a penalty? Second, what is the relevant response? At this point, it is important to note the scope of the analysis that this thesis will adopt in answering each question.

As to the causative event question, to date no one has contended that penalties arise in response to a miscellaneous 'other event'. This may be because there are no similarities between penalties and the events that live in the dustbin category (such as accessions, mixtures and specifications). In any event, an explanation of a right as arising from a non-miscellaneous event is more powerful than an explanation that links the right to a miscellaneous event, as the former strengthens legal coherence.[45] Consequently, this section passes over this remote possibility. It only considers the three non-miscellaneous causative events: wrongs, unjust enrichment and consent.

As to the responses question, the previous section identified the four legitimate responses that may characterise a right to liquidated damages. It also discounted a fifth. If a function legitimately may ground a right to liquidated damages, it cannot also ground a right to a penalty.[46] By process of elimination, a right to a penalty must display either a punitive function, or a miscellaneous 'other function'. No obvious miscellaneous function exists. At first blush, then, only one *plausible* option exists: the right to a penalty is best understood as operating punitively. So far, this thesis has shown only that this proposition is plausible. This section presents an argument substantiating that proposition.

I The causative event

(1) Wrongs

The most straightforward view is that penalties arise from a general law wrong. As Chapter 1 discussed, a wrong is an infringement of a right held by another; or, put differently, the infringement of a primary right that generates a secondary right. So, if the Penalty Rule did not exist, it might be said that where a contract includes an *excessive* agreed damages clause, upon B's breach A obtains a secondary right against B for the penalty sum. If this analysis is correct,

[45] Andrew Burrows, 'The Relationship between Property and Unjust Enrichment: Some Unresolved Issues' in Simone Degeling and James Edelman (eds), *Unjust Enrichment in Commercial Law* (Lawbook Company 2008) 339 fn 22.

[46] More specifically, the function cannot ground a right to the *excessive component* of a penalty.

it follows that A's right to a penalty arises from the wrong of B's breach of contract, but the parties' ability to decide the value of that right arises from consent. In other words, a penalty clause involves an ordinary, general law secondary right arising upon a primary right being infringed, though the parties determine the quantum of that secondary right.

Lord Diplock's comments in *Photo Production* seemingly support this interpretation. His Lordship stated that an agreement must not traverse the Penalty Rule: that is, 'it must not impose upon the breaker of a primary obligation *a general secondary obligation* to pay to the other party a sum of money that is manifestly [excessive] …'.[47]

Three points dispel this myth.

To start, consider again Lord Diplock's statement. His Lordship in fact held that it is *the agreement* that must not impose the impugned, 'secondary' obligation. It is not the general law. Accordingly, the statement in fact supports the view that the relevant right is consensual—and primary—in nature. It does not support the proposition that the right is generated by the general law wrong of breach of contract.

Next, Chapter 6 demonstrated that, as best understood, the way that A enforces her right against B under a penalty clause is by suing on the clause. Conversely, where a defendant commits the wrong of breach of contract, the claimant enforces her right against the defendant by bringing a general law claim for damages. The method of recovery, then, suggests that the right to a penalty is not a right that responds to a general law wrong.

Lastly, to propound this view is to commit a category error. It sees the causative event as the wrong, and the response as a consensual one (for example, a *response* of 'consensual punishment'). This cannot be. Consent is an event, not a response. It lines up on the horizontal axis of Table A, not the vertical one.[48] If consent is properly not the class of real world facts that engenders the response, it cannot re-enter the fray to influence the response. If that were possible, the taxonomy would be inconsistent. Explanatory force would be lost.

The conclusion is simple: the right to a penalty does not arise from a general law wrong.

(2) Unjust enrichment

[47] *Photo Production Ltd v Securicor Transport Ltd* [1980] AC 827 (HL) 850.

[48] Table following n 57 in ch 1.

Powerful voices have contended that the right to a penalty arises from an unjust enrichment. Examples from the bench and the academy exist. As to the former, in *Esanda Finance Corp Ltd v Plessnig*, Deane J held that relief from a penal provision might result from the 'principles of unjust enrichment operating in all the circumstances of the case'.[49] As to the latter, Bryan argues that the Penalty Rule exemplifies a 'kind of anticipatory unjust enrichment',[50] whereby in the paradigm scenario[51] the private law intervenes before B has satisfied his penal duty to pay money to A in order to *prevent* an unjust enrichment occurring. The doctrine is thus prophylactic in nature. Where B has paid over to A the penal sum, the prophylaxis has failed, and restitution for unjust enrichment responds in order to 'reinforce' the Penalty Rule's impugned protective rationale. In essence, Bryan contends that A's right to a penalty arises from an anticipated unjust enrichment, and B's satisfying of his counterpart duty brings that unjust enrichment to fruition.

Scholars supporting the unjust enrichment classification have expressed a view on two important issues: the nature of the enrichment, and the relevant unjust factor.

Starting with the enrichment, Collins propounds that the private law is here seeking to prevent 'unjust enrichment by overcompensation': that is, the Penalty Rule seeks to obviate a situation where a creditor is unjustly enriched by receiving 'a remedy which *far exceeds* his loss resulting from the debtor's breach of contract'.[52] Therefore, the writer suggests that the amount of this 'overcompensation' represents the relevant enrichment; and that, in the case of an agreed compensatory damages clause, there must be a large discrepancy between the agreed compensatory damages sum and the creditor's anticipated loss.

Moving to the unjust factor, Birks provided the leading theory.[53] The writer argued that relief against penalties arises due to unjust enrichment, based on the unjust factor of 'inequality'. The reasoning underpinning this unjust factor is as follows. Borrowers cannot rationally evaluate the risks they are undertaking, because they need credit and are systematically over-optimistic about the future. A debtor's eagerness to sacrifice her long-term prospects to short-term advantage leads to a particularised vulnerability that impairs her

---

[49] (1989) 166 CLR 131 (HCA) 155.

[50] Michael Bryan, 'Equitable Relief from Forfeiture: Performance or Restitution?' in Charles Rickett (ed), *Justifying Private Law Remedies* (Hart Publishing 2008) 364.

[51] Text following n 20 in ch 1.

[52] Hugh Collins, *The Law of Contract* (4th edn, LexisNexis 2003) 388-89 (emphasis added).

[53] cf Michael Bryan, 'Equitable Relief from Forfeiture: Performance or Restitution?' in Charles Rickett (ed), *Justifying Private Law Remedies* (Hart Publishing 2008) 379 (a policy-motivated unjust factor of 'preventing penalties').

judgment, and requires protection. In Birks's words, 'Sometimes personal happiness is too much at stake for ordinary wisdom to exercise any restraint.'[54] In the paradigm scenario,[55] B suffers from a more generalised manifestation of this particularised vulnerability, as B is willing to accede potentially to dire future consequences in return for a perceived immediate benefit. Accordingly, although otherwise autonomous, B is unable to protect his own best interests in the contracting process, and needs protection.

The intervening basis, then, is unashamedly paternalistic: due to his systematic over-optimism, inevitably B will lose his capacity to assess rationally and self-manage when faced with a transaction containing a penalty clause; this vulnerability means that the parties do not transact on an equal basis; and the private law invalidates the clause to protect B from this vulnerability and redress the inequality.

The unjust enrichment analysis is superficially attractive. However, after close scrutiny, it fails for at least the three reasons.

First, the unjust enrichment view fails to explain the very matter in issue. So, suppose that B has paid over to A the penal sum. If unjust enrichment is the causative event, it seemingly follows that the relevant unjust factor both invalidates the penalty clause *and* grounds B's restitutionary claim that follows. However, this analysis only works if the putative unjust factor renders the clause voidable or terminable, for only these forms of invalidity—and the power to avoid (and revest) or the power to terminate respectively—arise from an unjust enrichment.[56] Conversely, as stated, a true penalty clause is partially unenforceable, not voidable or terminable. Consequently, the invalidity does not arise from an unjust enrichment, and no 'unjust enrichment power' exists. Burrows's interpretation of the law here is better. The writer argues that the Penalty Rule is a 'contractual doctrine' that operates simply by invalidating penalty clauses; and that B obtains an unjust enrichment claim right only if the rule impugns the clause *and* if B satisfies the elements of an unjust enrichment claim, which requires that B has already paid over to A the penal sum.[57]

Accordingly, properly analysed, one must consider two questions. The first question is one of *contractual validity*: is the agreed damages clause valid? The second question concerns *consequential recovery*: if the clause is (partially) invalid, how does B recover from A the penal

---

[54] Peter Birks, *An Introduction to the Law of Restitution* (Clarendon Press 1985) 208-209.

[55] Text following n 20 in ch 1.

[56] Peter Birks, *Unjust Enrichment* (2nd edn, OUP 2005) 126.

[57] Andrew Burrows, *The Law of Restitution* (3rd edn, OUP 2011) 353-54. See similarly Graham Virgo, *The Principles of the Law of Restitution* (2nd edn, OUP 2006) 337.

element of the sum paid over? These two questions must be separated. In the penalty clause context, unjust enrichment can only answer the second question. Thus, where the penalty clause is executory, unjust enrichment is irrelevant. Conversely, where B has already satisfied his penal duty, unjust enrichment provides the mechanism by which A is compelled to return to B the excessive component.[58] A simple conclusion follows. A penalty clause creates a right to a penalty. Unjust enrichment cannot explain why the right to a penalty, and the clause which creates it, is partially invalid. If it cannot explain why the right to a penalty is partially invalid, then it necessarily cannot elucidate how that right is generated in the first place.

Second, the alleged 'enrichment' is problematic. Unjust enrichment theorists argue that the amount of any 'overcompensation' constitutes the enrichment; and that this amount must represent a *substantial* discrepancy between the agreed compensatory damages sum and A's prospective loss. However, on orthodox principles, there is no minimum amount by which a defendant can be unjustly enriched.[59] Accordingly, even a small degree of pecuniary divergence should always suffice to provide an enrichment. But we are told that it cannot.

Finally, in truth, no satisfactory unjust factor exists. For instance, Birks's theory that inequality grounds the unjust enrichment claim is defective. So, Birks was ambiguous as to whether this unjust factor was rightly 'defendant-sided' or 'claimant-sided'. Support for both views exists.[60] They are each considered below.

On the one hand, if the unjust factor is properly based on an *exploitation* of vulnerability, then it is defendant-sided. However, Birks later doubted whether any defendant-sided unjust factor could ever ground an unjust enrichment claim right.[61] The better view is that Birks's doubts were well founded. It cannot.

On the other hand, if the ground of restitution is based on the claimant's *impaired consent* arising from his vulnerability, then it is claimant-sided. However, this interpretation does not match how the private law works in this area. If the law were in fact concerned with responding to impaired consent, then a court would inquire into B's particular circumstances in each Penalty Rule case, relieving him if his judgment actually had been impaired. But this is not how Anglo-Australian private law identifies a penalty clause.

---

[58] See *Paciocco v Australia and New Zealand Banking Group Ltd* [2014] FCA 35, (2014) 309 ALR 249 [359], [363], [365]-[366]. The relevant unjust factor is said to be mistake.

[59] Presumably subject to the principle *de minimis non curat lex* (that is, 'The law takes no account of trifling matters').

[60] Peter Birks, *An Introduction to the Law of Restitution* (Clarendon Press 1985) 209, 211, 213 (defendant-sided) cf 140, 204 (claimant-sided).

[61] Andrew Burrows, *The Law of Restitution* (3rd edn, OUP 2011) 339.

In brief, unjust enrichment is not the causative event that generates a right to a penalty.

(3) Consent

The final plausible causative event is consent. On this view, consent is the event eliciting a primary right to a penalty, rather than the event that merely determines the quantum of a secondary right arising from a general law wrong. In other words, the parties' agreement—and not any external legal rule—is the source of B's obligation to pay the penalty. If the Penalty Rule did not exist, then, an enforceable right to a penalty would arise from consent.

Is this analysis sound? It is. Three reasons may be given.

To start, in *Photo Production*, Lord Diplock held that it was 'the agreement', and not the general law, that imposed upon B an impeachable obligation to pay a penal sum.[62] Similarly, Smith has noted that the Penalty Rule is properly 'part of the general law of contract'.[63] This statement suggests that the right to a penalty, and the limits placed upon that right, arise from the law of consent.

Next, as previously mentioned, the most significant classificatory feature of a debt is that it usually arises from consent. As Chapter 6 demonstrated, penalties are conceptually akin to a debt. Admittedly, at some points the analogy with a debt is unhappy. But the presently important point remains: penalties are more than in the nature of a debt, which is the product of consent, than in the nature of general damages, which are the product of a general law wrong.

Lastly, a consent-based analysis is logically unimpeachable. A right to liquidated damages arises by consent; a right to a penalty differs only in *degree*, not *kind*. Logic thus dictates that a right to a penalty arises by consent. The resulting picture is rational, coherent and attractive.

As with a right to liquidated damages, so here a consent-based event 'picks up' on the structure of a wrong: the waters of consent are again poured into a wrongs-shaped vase. Accordingly, although a right to a penalty is a *primary* right arising from *consent*, it will exhibit the same structure or anatomy as a general law *secondary* right that arises from a general law *wrong* of breach of contract. In this thesis, the (inelegant phrase) 'consent-based wrong' refers

---

[62] *Photo Production Ltd v Securicor Transport Ltd* [1980] AC 827 (HL) 850.

[63] Lionel Smith, 'Relief Against Forfeiture: a Restatement' (2001) 60 CLJ 178, 183.

to a consent-based event that takes the shape of a general law wrong. The right to a penalty arises from such an event.

Accordingly, although it appears that the right to a penalty arises from a general law wrong, in truth it does not. The right to a penalty is a primary right generated by consent. Consent is the only causative event in play, although a breach of contract does form part of the story.

This situation is not unique. Consider, for example, the award of an agreed sum. Where a defendant agrees to pay money as the price of goods sold or as remuneration for work done and later does not pay, a claimant may seek an award of an agreed sum to enforce the defendant's primary duty to perform. Routinely, the claimant will rely on the fact the defendant has breached his duty in seeking the order. However, the reason a court makes the award is not because the defendant has breached the contract, but rather because the breach of contract evidences that the defendant did not pay the sum of money when due. Although a court may grant the remedy *when* a defendant has breached his contract with the claimant, it is not granted *for* that breach of contract. The remedy granted *for* a breach of contract is general damages. In short, in making her claim in court the claimant stands on her primary right, not her secondary right, although a wrong forms part of the story.[64]

One further complication must be addressed. As Chapter 4 showed, as best interpreted the Penalty Rule jurisdiction is engaged were B commits an actual breach of contract or a substantive breach of contract, and the relevant event is the trigger for B's detriment. Thus, a right to a penalty may arise from *either* an actual breach of contract or a substantive breach of contract, and hence a substantive wrong. Notably, a substantive breach can only occur where the parties have attempted to draft around the rule. A substantive wrong can occur in no other context. This necessarily means that a substantive breach of contract can only ever arise from the parties' agreement, and their consent. In other words, a substantive wrong can only ever be consent-based. In summary, then, the right to a penalty may arise from a 'consent-based actual wrong', or 'consent-based substantive wrong'; that is, a consent-based default.

To conclude, the right to a penalty is a primary right that arises from a consent-based default. When it comes to penalties, everything happens in the land of consent.

---

[64] Peter Birks, 'Rights, Wrongs, and Remedies' (2000) 20 OJLS 1, 27-28.

II The response

By process of elimination, this thesis has established a prima face case that the right to a penalty is best understood as operating punitively. It follows that, given the Penalty Rule renders that right partially unenforceable, the rule implements a principle of preventing consensual punishment. The following section seeks to substantiate these propositions.

(1) Support

In *Andrews v Australia and New Zealand Banking Group Ltd*, Counsel for the successful applicants submitted that there is 'a deep root in the law for the proposition that courts will not allow the strict performance of voluntarily assumed private law obligations to extend into the field of one party punishing another for private gain'.[65] In other words, for over a century, Anglo-Australian private law has invalidated penalty clauses based on a principle which holds that consensual punishment is prohibited.

In England and in Australia, venerable authority supports the view that penalty clauses are punitive in nature, and that the private law prohibits their enforcement on that basis.

On the one hand, in England a litany of supporting statements exists. Take the following four examples. First, in *Thompson v Hudson*, Lord Westbury held that a penalty clause 'is a punishment, an infliction, for not doing, or for doing[,] something'.[66] Second, in *Robertson v Alexander (Driver's Trustees)*, Lord Young encapsulated the law as follows: if the penalty clause 'be truly a penalty—that is, a punishment—the Court will not allow that, because the law will not let people punish each other. They may [not] contract… for punishment.'[67] Third, in *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd*, Lord Parmoor referred to a penalty as being inserted 'as a punishment on the defaulter'.[68] Finally, in *Jobson v Johnson*, Dillon LJ emphasised that the Penalty Rule's basis is that 'parties cannot lawfully enter into an agreement that the one party shall be punished at the suit of the other'.[69]

On the other hand, the Australian High Court has similarly endorsed the view that the right to a penalty involves an impermissible punitive response. For instance, in *Legione v Hateley*, Mason and Deane JJ held that, 'A penalty, as its name suggests, is in the nature of a

---

65 [2012] HCA 30, (2012) 247 CLR 205, 206-207.

66 (1869-70) 4 LRHL 1 (HL) 28.

67 1881 SC 555 (IH) 562.

68 [1915] AC 79 (HL) 100.

69 [1989] 1 WLR 1026 (CA) 1033.

punishment for non-observance of a contractual stipulation …'.[70] Later, in *Acron Pacific Ltd v Offshore Oil NL*, Deane J described a penalty clause as imposing 'some additional … financial obligation in the nature of punishment'.[71] Recently, in *Andrews*, the High Court specifically adopted Mason and Deane JJ's statement in *Legione*, and cited with approval Deane J's statement in *Acron*.[72]

English and Australian academics have implicitly endorsed the proposition that penalty clauses are punitive in nature. For example, McKendrick has argued that a penalty clause's 'impermissible' function is 'to punish the party in breach'.[73] Further, scholars from each jurisdiction have noted the Penalty Rule's underlying preventing punishment principle. For instance, Rowan has argued that the rule 'constitutes a resounding rejection of deterrence and punishment as acceptable aims in the law of contract'.[74]

The judicial and academic position, then, is clear.

(2) Relief co-extensive with the response

Chapter 6 dealt with the consequences of the Penalty Rule invalidating an agreed damages clause. It was shown that, if a penalty exists, Anglo-Australian courts adopt the 'compression' approach: that is, the penalty clause is compressed to the extent of its enforceability. Two points necessarily follow. For one, courts refuse to allow A to enforce against B the excessive component of the clause. In other words, B is not liable for the penal element. For another, they permit A to enforce against B the remaining, legitimate component of the clause. The preventing punishment principle explains these consequences: the relief is merely co-extensive with the response. This statement needs unpacking. Three points are necessary.

To start, a taxonomical study of the private law suggests that any relief will seek to implement a particular response on the extended Birksian grid, being Table A.[75] To implement the response satisfactorily, the relief should be *co-extensive* with the response. For example, in 'core cases' of unjust enrichment, the relief is an award of money had and received. This

---

[70] (1983) 152 CLR 406 (HCA) 445.

[71] (1985) 157 CLR 514 (HCA) 520.

[72] *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [9]-[10].

[73] Ewan McKendrick, *Contract Law* (10th edn, Palgrave Macmillan 2013) [21.5]. See similarly Elisabeth Peden, 'Forfeiture of Deposits: Where Law and Equity Collide?' (2012) 6 J Eq 161, 174.

[74] Solene Rowan, 'Reflections on the Introduction of Punitive Damages for Breach of Contract' (2010) 30 OJLS 495, 508. See similarly JW Carter and Elisabeth Peden, 'A Good Faith Perspective on Liquidated Damages' (2007) 23 JCL 157, 161.

[75] Table following n 57 in ch 1.

precisely implements the restitutionary response by requiring the defendant to 'give back' to the claimant the exact value of the enrichment unjustly obtained at her expense.

Next, Birks postulated that the law contained an 'intelligible policy requiring the elimination of all penal sanctions from the … private law'.[76] The writer suggested that the Penalty Rule pursued that objective, and that courts related relief to that objective. Put differently, the Penalty Rule seeks to eliminate punishment, and nothing more. The relief that implements that objective most perfectly will merely invalidate the clause's punitive component, and nothing more. The English Court of Appeal has recognised this.[77] Thus, the preventing punishment principle explains why courts do not enforce the clause's penal element. This negative principle requires that any agreed damages sum does not contain a punitive component, and the penal element is punitive in nature.

Lastly, it follows from the previous point that the preventing punishment principle explains why courts enforce a penalty clause's 'non-penal' element. Once the court has removed the penal element, there is no reason to stop A enforcing against B the remaining, legitimate component of the clause. So, if the court is dealing with an invalid agreed *compensatory* damages clause, the relief must enforce the clause's valid compensatory component. As Heydon JA has noted, the private law 'ensures that contract breakers are not punished by penalties, though it also ensures that even though the penalty clause is rendered unenforceable, [incurred] losses are recovered'.[78]

In conclusion, a classificatory analysis demonstrates that, on the extended Birksian grid, the right to a penalty is properly placed in the box where consent and punishment intersect. But in Anglo-Australian private law, penalty clauses are banned. Therefore, the Penalty Rule demands that this particular box in Table A has no content. Notably, this is not the only box that contains no content.[79] As Chapter 1 explained, this phenomenon is unproblematic, for the grid is intended merely to ask questions and identify possibilities. However, as McInnes has noted, in some cases Birks's taxonomy 'helps to eliminate responses that ought not to exist'.[80] Punishment is precisely that.

---

[76] Peter Birks, *An Introduction to the Law of Restitution* (Clarendon Press 1985) 214.

[77] *Oresundsvarvet AB v Lemos (The Angelic Star)* [1988] 1 Lloyd's Rep 122 (CA) 127.

[78] *Harris v Digital Pulse Pty Ltd* [2003] NSWCA 10, (2003) 56 NSWLR 298 [338].

[79] For instance, the only response to unjust enrichment is restitution, such that the other responses fail to arise. Consequently, other boxes—such as that where unjust enrichment and punishment intersect—have no content.

[80] Mitchell McInnes, 'Taxonomic Lessons for the Supreme Court of Canada' in Charles Rickett and Ross Grantham (eds), *Structure and Justification in Private Law* (Hart Publishing 2008) 89.

III Consequences

Correct classification has 'important consequences'.[81] A categorical study of liquidated damages clauses and penalty clauses reveals three.

(1) Proscriptive, not prescriptive

A taxonomic investigation shows that the Penalty Rule is not *prescriptive* in nature. Contrary to orthodox views,[82] properly understood the rule does not mandate that all agreed damages clauses implement a compensatory response. Put differently, it is not essential that a right to liquidated damages is compensatory in nature. It follows that courts legitimately may enforce agreed damages clauses that embody other responses, such as substitution and disgorgement.

Rather, the Penalty Rule is *proscriptive* in nature. As best interpreted, the doctrine prohibits any agreed damages clause that operates punitively. In other words, a party cannot recover any agreed damages amount that 'goes beyond' a sum that instantiates a legitimate response: that is, A may not recover from B an amount that is 'over and above' a substitutive, compensatory, restitutionary or profit stripping sum. These parasitic, excessive sums are punishment exemplified.

(2) False dichotomy

A further taxonomic insight reveals a fundamental error in modern judicial thinking. It is first necessary to set out the background. The misstep will then be exposed.

The story begins with *Lordsvale Finance plc v Bank of Zambia*.[83] In that case, Colman J sought to interpret the 'central' case of *Dunlop* and provide a gloss on the test of validity. The judge held that to identify a penalty clause a court must ask whether, as at the contract date, the clause's predominant function was 'to deter [B] from breaking the contract or to compensate [A] for breach'.[84] In other words, a dichotomy exists between agreed damages clauses that are compensatory, and those that are deterrent. The former are valid; the latter are not.

[81] *Foskett v McKeown* [2001] 1 AC 102 (HL) 129.

[82] This assumption underpins the conventional preventing 'overcompensation' theory, discussed in chapter 7.

[83] [1996] QB 752 (QB).

[84] *Lordsvale Finance plc v Bank of Zambia* [1996] QB 752 (QB) 762.

Deterrence may take two forms. It may be specific, which is aimed at the individual; or general, which is aimed at society as a whole. The premise underlying this dichotomy is that an agreed damages clause seeking to deter B *specifically* from breaching the contract is invalid.

English appellate courts have enthusiastically supported this dichotomy. For example, in *Cine Bes Filmcilik ve Yapimcilik AS v United International Pictures*, Mance LJ held that Colman J's dichotomy provided an 'accessible paraphrase of the concept of penalty'.[85] Recently, in *Makdessi v Cavendish Square Holdings BV*, the English Court of Appeal supported the dichotomy, stating that a provision is penal if its function is 'not to compensate but to deter breaches of obligations …'.[86] The case law is all one way. The contagion has also spread to Australia.[87]

However, the dichotomy between compensation and deterrence exhibits a severe collapse in legal understanding. It displays a theoretical and classificatory error, and is seriously inadequate. Each issue is considered in sequence.

First, this dichotomy contains a theoretical flaw: deterring breach is merely the flipside of incentivising performance, which is legitimate. To explain, *any* agreed damages clause will incentivise B to perform his contract with A. The clause will also deter B from breaching the contract to precisely the same degree. This is because the two are simply flipsides of the same coin. In *Makdessi*, the Court of Appeal acknowledged this, stating that the instant clause was penal because 'its function [was] to deter breach or, to put it positively, to secure performance …'.[88] As there is no meaningful distinction between incentivisation and deterrence, the proposed dichotomy cannot provide a sound basis for distinguishing between valid and invalid agreed damages clauses.

Second, this alleged dichotomy between compensation and deterrence conceals a grave category error. As Chapter 1 noted, the Birksian taxonomy adopts categories that are mutually exclusive. It does so to avoid overlaps. However, according to the logic of this dichotomy, deterrence is properly a category of response. Accordingly, it should form a header in the 'Responses' rows of Table A, and be situated alongside substitution, compensation, restitution, disgorgement, punishment, perfection and 'other functions'. However, this would create several overlaps: deterrence cuts across all the non-miscellaneous categories, and possibly even traverses the miscellaneous one.

---

[85] [2003] EWCA Civ 1669, [2004] 1 CLC 401 [13].

[86] [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [125].

[87] *Fermiscan v James* [2009] NSWCA 355, (2009) 261 ALR 408 [134].

[88] *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [125].

Consider again the proposed dichotomy of compensation versus deterrence. These two concepts are not intrinsically discrete, as the prospect of having to pay compensation will itself involve some deterrence. As Popplewell J noted in *Imam-Sadeque v Bluebay Asset Management (Services) Ltd,* 'Any clause which provides for payment of money on breach is likely to operate to some extent as a deterrent against breach, even if it fulfils a compensatory function.'[89]

Why does deterrence create overlaps? This is because deterrence does not line up on the same taxonomic plane as responses such as compensation and disgorgement. In truth, deterrence is not a category of response; rather, it is an *effect* of a right that falls within such a category. Properly understood, deterrence is not a function; it is an effect, and most forms of relief will have a deterrent effect. For instance, a general compensatory damages award will itself deter. But the fact that general compensatory damages will have a deterrent effect on the defendant (or like defendants in future) does not mean their purpose is to deter—the deterrent impact is merely an incidental by-product of the right's true function, which is to compensate for loss. Deterrence, then, is a useful *consequence of* liability, but not a *reason for* it.[90] The fact that the private law has coincidental effects should not lead us to conclude that these effects are its functions. That would be a big mistake.

Compare another dichotomy: compensation and *punishment*. Unlike Colman J's juxtaposition, this dichotomy contrasts functions that are mutually exclusive. The two notions are conceptually distinct. By definition, a punitive payment does not seek to compensate. Accordingly, compensation versus punishment is the true dichotomy. Much support exists for this juxtaposition. For example, in *Ruxley Electronics and Construction Ltd v Forsyth*, Lord Lloyd held that 'the object of damages is always to compensate the [claimant], not to punish the defendant', for damages 'are compensatory not punitive'.[91] Equally, several Australian High Court justices have recognised this contrast, commenting that penalty provisions are those whose nature or character is 'penal rather than compensatory'.[92]

Notably, judicial comments regarding deterrence are explicable in terms of this dichotomy. So, properly understood, the Penalty Rule *does* prohibit the specifically deterrent

---

89 [2012] EWHC 3511 (QB), [2013] IRLR 344 [189].

90 See similarly Ernest J Weinrib, 'Deterrence and Corrective Justice' (2002) 50 UCLA L Rev 621, 624-25, 633, 638-39.

91 [1996] AC 344 (HL) 365, 373.

92 *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 193; *Esanda Finance Corp Ltd v Plessnig* (1989) 166 CLR 131 (HCA) 138.

effect that a penalty clause produces. However, this is not because the rule is interested in that effect *per se*. Rather, it is because the specifically deterrent effect flows from a prohibited punitive means. The deterrent effect is, as it were, 'fruit from the poisonous tree'. For instance, in *Andrews*, Gordon J held that the court must determine if the agreed damages sum is payable 'as a punishment to deter breach of the contractual obligation …'.[93] Similarly, in *Jeancharm Ltd v Barnet Football Club Ltd*, Keene LJ held that a penalty clause existed 'to deter the party in question from breaking the contract by providing for a punitive level of payment'.[94] Deterrence, then, is only impermissible where punishment is used as the instrument to deter. It follows that it is permissible for a party to seek to achieve a specifically deterrent effect by another means. This is indeed the case. Assume that, in the paradigm scenario,[95] A demands that her contract with B include an agreed compensatory damages clause stating that B must pay to A the stipulated sum if he breaches the contract *on a Wednesday*. A clearly includes this clause specifically to deter B's breach of contract, though only if it occurs on a Wednesday. Despite this, if the clause passes the test of validity, it is enforceable.

Finally, the compensation versus deterrence dichotomy leaves no room for legitimate, non-compensatory responses. It is therefore inadequate. As this thesis has demonstrated, Anglo-Australian case law supports the view that a right to liquidated damages may function by substituting for the value of an infringed right, or by divesting a profit. Accordingly, though a correct dichotomy *must* include punishment, it need not necessarily include compensation. Thus, although one correct dichotomy is compensation versus punishment, *several* other proper dichotomies exist. For instance, another valid juxtaposition is between substitution and punishment. Notably, English judges have recognised this. In *Ruxley*, Lord Bridge contrasted two aspects of damages for breach of contract: they may reflect the fact that the claimant 'did not get what [she] bargained for', but they must raise no prospect of 'punishing the contract breaker'.[96]

Put simply, the dichotomy between compensation and deterrence is false. It is a chapter of the Penalty Rule story that is best forgotten.

(3) A punitive damages analogue

---

[93] *Andrews v Australian and New Zealand Banking Group Ltd* [2011] FCA 1376, (2011) 211 FCR 53 [58], revd [2012] HCA 30, (2012) 247 CLR 205.

[94] [2003] EWCA Civ 58 [22].

[95] Text following n 20 in ch 1.

[96] *Ruxley Electronics and Construction Ltd v Forsyth* [1996] AC 344 (HL) 353.

The final point to arise from the classificatory theoretical analysis is crucial. Given the taxonomic nature of a right to liquidated damages and a right to a penalty, and the relationship between each type of right and a right a general damages award, two points arise. First, liquidated compensatory damages are a general compensatory damages analogue. Second, penalties are a punitive damages analogue.

To make sense of these statements, one must accept that the private law adopts the traditional *associated* view.[97] So, Chapter 5 discussed two approaches to agreed damages clauses: an associated view and a disassociated view. On the associated view, the general damages and agreed damages regimes are interconnected. In particular, the agreed damages regime is merely an *analogue* of the general damages regime. It follows that, under any agreed damages clause, A may only recover from B an amount that suitably reflects the amount that a similarly situated claimant would recover from a defendant pursuant to a general law remedy. As Burrows has contended, the validity of an agreed damages clause will be 'assessed in line with the judicial remedies available. The further the parties stray from what the courts would themselves award, the more likely it is that the agreed remedy will be regarded as penal …'.[98] Similarly, McGhee has suggested that courts will only uphold an agreed disgorgement damages clause where it properly reflects the general law position regarding a profit-stripping award.[99]

The associated view inexorably leads to the conclusion that, in the 'core case', liquidated compensatory damages are a general compensatory damages analogue. Given this, another irresistible analogy from the law of wrongs beckons. Properly understood, penalty clauses are a punitive damages analogue. If that is correct, then the principle proscribing one form of remedy is analogous to the principle proscribing the other. That is, the Penalty Rule's rationale is one of preventing consensual punishment, whilst the logic behind the general rule against punitive damages is one of preventing imposed punishment.

The view that a right to a penalty and a right to punitive damages are analogous rights has much support. For example, Burrows argues that the most straightforward and strongest argument for prohibiting penalty clauses is one based upon an analogy between these clauses and punitive damages: namely, that enforcing a penalty clause would 'contradict the traditional approach to damages according to which punitive damages are not available for breach of

---

[97] As Chapter 7 noted, the orthodox preventing 'overcompensation' theory takes this view.

[98] Andrew Burrows, *Remedies for Torts and Breach of Contract* (3rd edn, OUP 2004) 445.

[99] John McGhee (gen ed), *Snell's Equity* (33rd edn, Sweet and Maxwell 2014) [13-013].

contract'.[100] Similarly, many contend that a like principle prohibits the right to a penalty and the right to punitive damages. For instance, Rowan asserts that the same principle underlies both the Penalty Rule, and the rule against punitive damages for breach of contract. The former constitutes a 'resounding rejection' of punishment as an acceptable response; the latter spurns the punitive motivation that lies 'at the core of punitive damages'.[101]

It has been argued that 'A state-supplied right to sue for punitive damages is similar to a contract-supplied right to sue for a penalty'.[102] That proposition is undeniable. The more difficult question is how many normatively important similarities exist. The answer is two.

For one, these two rights share the same form. A punitive damages sum involves an amount that is parasitic on, and goes 'above and beyond', any amount required to satisfy another non-miscellaneous response (such as compensation). The same applies to the excessive component of the penal sum.

For another, it follows that the right to punitive damages and the right to a penalty share the same function: punishment. As to the former, it is widely recognised that the role of punitive damages is to punish the defendant.[103] As to the latter, the excessive component of an agreed damages sum—the penal sum—operates to punish B.

Judges have long noted the parasitic nature of punishment in each context. As regards punitive damages, in *Cassell and Co Ltd v Broome*, Lord Reid drew a distinction between compensatory damages and 'damages which go beyond [such damages] and are purely punitive'.[104] In short, punitive damages are always parasitically attached to another form of damages. As regards penalty clauses, recall Spieglman CJ's hypothetical example in *Harris*: his Honour postulated an agreed damages clause stating that 'the employer could recover, for a relevant breach, either [compensatory] damages or [disgorgement damages] plus $10,000. It is clear that the … $10,000 would be struck down as a penalty.'[105] Issues of quantification aside, the statement clearly illustrates that courts conceptualise the punitive component of a penalty amount as being inherently parasitic upon an amount that implements another response, such as compensation or disgorgement.

---

[100] Andrew Burrows, *Remedies for Torts and Breach of Contract* (3rd edn, OUP 2004) 451.

[101] Solene Rowan, 'Reflections on the Introduction of Punitive Damages for Breach of Contract' (2010) 30 OJLS 495, 497, 508.

[102] Alan Schwartz, 'The Myth that Promisees Prefer Supracompensatory Remedies' (1990) 100 Yale LJ 369, 370.

[103] eg *Rookes v Barnard* [1964] AC 1129 (HL) 1221.

[104] [1972] AC 1027 (HL) 1086.

[105] *Harris v Digital Pulse Pty Ltd* [2003] NSWCA 10, (2003) 56 NSWLR 298 [58].

If these are the relevant similarities, are there any relevant differences? The causative event differs: a right to punitive damages would arise from a general law wrong, whereas a right to a penalty would arise from consent. In the latter case, it is the parties—not the court—that dispense the punishment. Also, the assessment date is dissimilar: the quantum of a right to punitive damages is assessed *ex post*, while the quantum of a right to a penalty is determined *ex ante*. Two other potential differences, however, must be considered.

On the one hand, it has been said that one 'object' of punitive damages is to deter.[106] This statement is prone to mislead. As previously demonstrated, deterrence is not a category of response; rather, it is an effect of a right that falls within such a category. Thus, correctly analysed, courts do not award punitive damages for the *purpose* of deterring, but if they make such an award it will have a deterrent *effect*. Further, this statement obscures the nature of the deterrent effect that punitive damages produce. The Australian High Court has held that punitive damages awards are valuable because they produce a specifically *and generally* deterrent effect.[107] Conversely, Anglo-Australian courts consider that a penalty clause will produce a specifically deterrent effect, not a general one. Therefore, a right to punitive damages will produce a generally deterrent effect; a right to a penalty will not. This distinction is crucial.

On the other hand, some judges take the view that punitive damages perform a 'vindication' function: that is, they 'assuage any urge for revenge'.[108] This would provide a valid distinction with penalty clauses. Though superficially attractive, this view of the nature of punitive damages is wrong. As a matter of principle, punitive damages are not vindicatory. Vindication is about *the claimant*. Conversely, punitive damages are wholly focused upon *the defendant's* wrongdoing. Consequently, they totally disregard the claimant. As Beever notes, a court may decide that a defendant deserves to be punished even though it cares nothing for the victim of the wrong, or even if no victim exists at all. The writer concludes that, in awarding punitive damages, a court is 'expressing condemnation of the defendant, but condemnation of the defendant does not imply vindication of the claimant'.[109]

The short point is this. Penalty clauses are a punitive damages analogue, and the law prohibits each remedy based on a comparable preventing punishment rationale.

---

106 *Rookes v Barnard* [1964] AC 1129 (HL) 1221.

107 *Lamb v Cotogno* (1987) 164 CLR 1 (HCA) 10.

108 *Harris v Digital Pulse Pty Ltd* [2003] NSWCA 10, (2003) 56 NSWLR 298 [55].

109 Allan Beever, 'The Structure of Aggravated and Exemplary Damages' (2003) 23 OJLS 87, 99.

D Assessment

This chapter has thus far presented a classificatory theory of the Penalty Rule. It has demonstrated that, viewed taxonomically, the right to a penalty is a consensual punishment; and the Penalty Rule's true underlying principle is one of preventing that consensual civil punishment. Naturally, the question arises as to whether preventing consensual punishment is in fact justifiable. The remainder of this chapter attempts to answer that question.

It is essential to take stock. If one accepts the argument to this point, the second aim of this thesis—set out in Chapter 1—is achieved: it has set the terms for a proper jurisprudential debate concerning the rule's continued existence. Penalties represent a consensual punishment; and the Penalty Rule prohibits this. The proper issue to debate is whether this consensual punishment is, in fact, unjustifiable. All are now on the same page.

One returns, then, to Birks. The writer argued that preventing punishment is a 'low level' principle requiring 'its own higher justification'.[110] Therefore, any theory of the rule based on this principle is a grounded one. Accordingly, a satisfactory theory of the rule must go deeper, and present a more profound justification. As the following illustrates, that is precisely what corrective justice provides.

2 Corrective justice theory

This section outlines the corrective justice theory that justifies the prohibitive principle underlying the Penalty Rule. It first outlines the core, abstract principles comprising the corrective justice norm. The section then shows how the corrective justice norm relates to, and justifies, the principle against *imposed* punishment, and the analogous principle against *consensual* punishment. Lastly, it demonstrates that there is a further reason supporting the banning of consensual punishment that applies only to it, and does not apply to imposed punishment. Thus, corrective justice's prohibition on consensual punishment is justified 'by a stronger reason'.

A Corrective justice in a nutshell

In what follows, this thesis provides a basic account of Weinribian corrective justice theory. Admittedly, the account may strike those familiar with the voluminous correctivist literature as simplistic. However, this thesis only elucidates corrective justice theory to the extent necessary

---

110 Peter Birks, *An Introduction to the Law of Restitution* (Clarendon Press 1985) 214.

to justify the Penalty Rule. To do so, it suffices merely to grasp the foundational concepts constituting Weinrib's corrective justice theory. In short, the following account is fit for purpose, and for present purposes of satisfactory quality.

Aristotle famously distinguished between two norms of justice: distributive justice and corrective justice.

'Distributive justice' is the form of justice concerned with the overall distribution of a divisible good, such as wealth, among the members of a community. Distributive justice divides the good in accordance with some criterion that compares the member's relative merits. There are many criteria for assessing the justice of any distribution: for example, need, merit or skill. So, distributive justice embodies a 'proportional equality'.[111]

'Corrective justice' is the form of justice concerned with rectifying an injustice that has arisen in an interaction between individuals. In Weinrib's words, 'Corrective justice is the idea that liability rectifies the injustice inflicted by one person on another.'[112] This form of justice views interactions between two individuals through a close-up shot that focuses solely on those two individuals, rather than through a panoramic that views individuals in terms of their relationship with the members of a larger community. In essence, corrective justice demands fairness as between the parties.

For centuries, philosophers have debated the precise nature of the corrective justice norm. It is not necessary, nor desirable, for this thesis to contribute to that literature. Suffice it to say that, as Weinrib conceptualises it, corrective justice theory arguably displays five central features. These principles underpin Weinrib's overarching corrective justice norm. We examine these now.

I Bipolarity

The first feature is the 'bipolarity principle'. Taking a formalist approach,[113] this principle holds that the private law's most essential characteristic is structural: liability creates a 'direct nexus' between the claimant and the defendant which renders the parties' relationship bipolar in form.[114] Pursuant to this principle, a private law relationship is necessarily bipolar in that it links only two parties, and no more, as a finding of liability always connects a *particular* claimant

[111] Ernest J Weinrib, 'Corrective Justice in a Nutshell' (2002) 52 UTLJ 349, 349.

[112] Ernest J Weinrib, 'Corrective Justice in a Nutshell' (2002) 52 UTLJ 349, 349.

[113] Text following n 37 in ch 1.

[114] Ernest J Weinrib, *The Idea of Private Law* (Harvard UP 1995) 1-2, 10, 43, 76.

with a *particular* defendant.[115] Both parties, then, are linked like the two poles of an abacus. The bipolarity principle explains why the causation doctrine plays a fundamental role in determining liability in private law litigation.

II Injustice

The second feature is the 'injustice principle'. The basic idea behind this principle is as follows. Pursuant to corrective justice theory, two parties interact under an initial, notional equality or equilibrium. However, an event may occur which disturbs that equilibrium, causing one party to realise a gain and the other party to suffer a corresponding loss. This principle holds that as such an event has disrupted a pre-existing equilibrium between two individuals, it is properly labelled an 'injustice'.[116]

To upset the prior equilibrium and create an injustice, a defendant must be responsible for an 'inconsistency' with the claimant's right.[117] The prime example of such an inconsistency is where the defendant infringes the claimant's primary right, such as by breaching a contractual, equitable or tortious duty owed to the claimant. Thus, a wrong provides the 'core case' of an injustice. However, an event that is not a wrong may still involve an inconsistency with the claimant's right for which the defendant is responsible. For instance, many agree that an unjust enrichment event sufficient to create a claim right to restitution—such as the innocent receipt of a mistakenly transferred payment—constitutes an injustice.[118] As discussed below, corrective justice demands that injustices are corrected. The important point to note is that not all injustices requiring correction are rights violations grounding secondary rights. The orthodox view is that the corrective justice norm underpins and justifies recovery in the law of unjust enrichment,[119] but a claim right in unjust enrichment is a *primary* right.

III Correction

The third feature is the 'correction principle'. This principle requires that the relevant injustice be 'corrected' by removing the gain from one party and restoring it to other who has suffered

---

[115] Ernest J Weinrib, 'Corrective Justice in a Nutshell' (2002) 52 UTLJ 349, 351; Ernest J Weinrib, *Corrective Justice* (OUP 2012) 1, 338.

[116] Ernest J Weinrib, 'Corrective Justice in a Nutshell' (2002) 52 UTLJ 349, 349; Ernest J Weinrib, *Corrective Justice* (OUP 2012) 82.

[117] Ernest J Weinrib, *Corrective Justice* (OUP 2012) 88, 90, 116.

[118] eg Ernest J Weinrib, 'Corrective Justice in a Nutshell' (2002) 52 UTLJ 349, 353; Lionel Smith, 'Restitution: The Heart of Corrective Justice' (2001) 79 Tex L Rev 2115, 2141-42.

[119] Andrew Burrows, *The Law of Restitution* (3rd edn, OUP 2011) 68-69, 621-22.

the corresponding loss. This is said to remedy the notional imbalance in the parties' positions, and reinstate the pre-existing equilibrium. So, assume that X commits a wrong against Y, and notionally takes 10 units from Y. Consequently, 10 beads will slide towards X's end of the abacus. As X has violated a corrective justice norm, it requires that X restore the units to Y to correct the disequilibrium. Thus, 10 beads must be slid to Y's end of the abacus. Further, under this principle, the judicial relief undertakes the correction. It 'undoes' the injustice by forcing the party who has gained from the injustice to make good the position of the party who has suffered from it, thus restoring the parities' initial equality.

According to Weinrib, the remedy undertaking the correction is 'equivalent to' and 'limited by' the injured right.[120] The 'equivalency thesis' and 'limitation thesis' are crucial to understanding the corrective justice norm. We consider each in order.

(1) Equivalency

Beginning with the equivalency thesis, the correction principle requires that the remedy seek to undo the injustice in so far as it is possible to do so. According to Stevens, when dealing with a wrong, corrective justice undoes the injustice (as far as it is possible) by instating the 'next best' position. So, if a defendant infringes a claimant's primary right, the corrective justice norm requires that the resulting injustice 'be corrected in the way which is the next best position now achievable to the wrong not having occurred'.[121] Thus, the only correctively just response to a wrong is to place the claimant into the next best position, for this is the way in which the private law notionally undoes the wrong.

The concept of the next best position must be further elucidated. According to this idea, once a defendant has committed a wrong, the 'best' thing to occur would be for the defendant to comply with his original primary duty. But the time for performance has past. The clock cannot be turned back, and the rights infringement cannot literally be reversed. As such, the defendant's primary duty is replaced by a duty to do the 'next best' thing: that is, a duty to achieve 'the closest to full satisfaction that is still available'.[122] In other words, the defendant must do the next best thing to perfect compliance that is possible. Therefore, where a wrong has been committed, 'the secondary obligation to pay money imposed upon the

---

[120] Ernest J Weinrib, *Corrective Justice* (OUP 2012) 84.

[121] Robert Stevens, 'Rights and Other Things' in Donal Nolan and Andrew Robertson (eds), *Rights and Private Law* (Hart Publishing 2012) 145.

[122] John Gardner, 'What is Tort Law For? Part 1: the Place of Corrective Justice' (2011) 30 Law & Phil 1, 33.

wrongdoer [is] the law's attempt to reach the "next best" position to the wrong not having been committed … in the first place'.[123]

A secondary duty to pay general damages, then, is an obligation to do the next best thing to the primary duty that was not performed. However, this secondary duty comes in many forms, depending on the response in play. Consequently, the secondary duty representing the next best position may differ depending on the facts and the nature of the claimant's claim. For instance, if the claimant seeks to recover the value of an infringed right, a secondary right to general *substitutive* damages will constitute the closest position the law can achieve to the wrong not having occurred at all. Equally, if the claimant seeks to recover her consequential loss, then a secondary right to general *compensatory* damages is the next best thing that a court can award.

What is the next best position in practice?[124] Generally speaking, *all forms* of general damages awards for *all types* of wrongs attempt to put the claimant in the position she would have occupied had the defendant not wronged the claimant, in so far as money can do this. In all cases, this is what a secondary duty to pay damages seeks to do, and it represents the next best thing to the defendant not having violated the claimant's right in the first place. Specifically, where the wrong is a breach of contract, general damages seek to place the claimant in the situation in which she would have been if the defendant had performed his contractual obligations, to the extent that money can do this. In the contractual context, this constitutes the next best position to the defendant having in fact performed.

Much support exists for the idea that general damages awards seek to reach the next best position to the wrong not having been committed. For instance, in *Robinson v Harman*, Parke B held that if the claimant seeks general damages for breach of contract, she is 'so far as money can do it, to be placed in the same situation … as if the contract had been performed'.[125]

Importantly, the secondary right that seeks to achieve the next best position must bear a particular shape. This is due to its relationship with the primary right. Clearly, the secondary right is not the same as the primary right. Rather, the secondary right is 'the form the primary

---

[123] Robert Stevens, *Torts and Rights* (OUP 2007) 59 (footnote omitted).

[124] See Robert Stevens, 'Damages and the Right to Performance: a Golden Victory or Not?' in Jason W Neyers, Richard Bronaugh and Stephen GA Pitel (eds), *Exploring Contract Law* (Hart Publishing 2009) 175.

[125] (1848) 1 Exch 850, 855; 154 ER 363, 365. See also *Tabcorp Holdings Ltd v Bowen Investments Pty Ltd* [2009] HCA 8, (2009) 236 CLR 272 [13].

right takes after the violation'.[126] This thinking is not new. Long ago, Blackstone argued that a remedy for a wrong seeks to 'redress the party injured, by either restoring [her] right, if possible; or by giving [her] an equivalent'.[127] From this, Weinrib explains that the corrective justice norm requires that the defendant 'restore' the claimant's primary right. However, the various kinds of damages—substitutive, compensatory, and gain-based—each constitute a 'quantitative form of restoration' that restores to the claimant the 'monetary equivalent' of that right.[128] In short, a secondary right to general damages is the monetary equivalent, or the monetarised version, of the primary right that the claimant held at the beginning of the story. It follows that, in the case of a breach of contract, a general damages award is simply the form that the primary obligation to perform takes after breach. Accordingly, the secondary right to general damages for breach of contract is merely a financial quantification of the primary right to obtain performance.

(2) Limitation

Turning to the limitation thesis, the correction principle requires that the remedy 'restores *only* the [claimant's] right and does not give the [claimant] more than that right (or its equivalent). Thus, the reason for creating liability also limits it.'[129] Put differently, a remedy must restore in full the right or its equivalent, but the corrective justice norm prohibits it going any further. Any additional duty is a sanction, and sanctions are a matter for the criminal law, not for the private law. Consider a secondary right to 'multiple damages'. Such damages require a defendant to pay an amount that is a multiple of the sum needed to compensate the claimant. A secondary right to multiple damages would give the claimant more than the equivalent of her right. Properly understood, then, multiple damages are 'a matter of public law', only potentially justifiable by 'some penal policy'.[130]

IV Monism

The fourth feature is the 'monism principle'. As Chapter 1 discussed, on a monist view the right and the remedy are perfectly congruent: the remedy is 'the same thing as the right, looked

---

126 Robert Stevens, 'Rights and Other Things' in Donal Nolan and Andrew Robertson (eds), *Rights and Private Law* (Hart Publishing 2012) 136.

127 4 Bl Comm 7.

128 Ernest J Weinrib, *Corrective Justice* (OUP 2012) 94-95, 172.

129 Ernest J Weinrib, *Corrective Justice* (OUP 2012) 92 (original emphasis).

130 Lionel Smith, 'Relief Against Forfeiture: a Restatement' (2001) 60 CLJ 178, 188 (footnote omitted).

at from the other end'.[131] Pursuant to the monism principle, a claimant's right is said to survive the injustice and continue into the remedy with a changed composition: the remedy is thus the shape that the right takes after the injustice. On this view, the right and remedy form 'a single continuous normative sequence'; and although the right may metamorphose throughout that sequence, its essence remains such that 'the right both grounds and limits the remedy'.[132] So, although the form may change, the substance remains the same. As Weinrib has argued, 'When one gazes at the right from the other end, one looks at the same thing even though what one may see is different.'[133] In other words, the right is the chrysalis, and the remedy is the butterfly that later emerges.

Therefore, there is a link between the monism principle and the equivalency thesis: the latter logically follows from the former. Under the monism principle, the secondary right (or remedy) 'fits' the primary right because the secondary right is a *continuation* of the primary right. However, the secondary right is not identical to the primary right; rather, the secondary right is the primary right in a *new or different form*. As the equivalency thesis notes, that new form must constitute a financial equivalent of the primary right. As discussed, the equivalent secondary right is a right to general damages, which right will differ according to the relevant response. Each kind of secondary right represents an instantiation of the anterior primary right. The secondary right to general *compensatory* damages is the most common. But it does not have a monopoly.

V Fairness

The fifth feature is the 'fairness principle'. A key aspect of corrective justice is the idea that the parties interact with each other as free and equal agents. To affirm and maintain the parties' freedom and equality, the private law must treat the parties fairly, which requires that it does not subordinate one party's interests to those of the other. This notion leads to the fairness principle. This principle holds that private law adjudication must be exclusively concerned with *the parties' relationship*, rather than concerned with each party taken individually. As Weinrib notes, corrective justice demands that all normative considerations that apply to the parties' relationship must 'reflect the parties' correlative standing as the sufferer and doer of

131 Peter Birks, 'Definition and Division: A Meditation on Institutes 3.13' in Peter Birks (ed), *The Classification of Obligations* (Clarendon Press 1997) 24.

132 Ernest J Weinrib, *Corrective Justice* (OUP 2012) 109.

133 Ernest J Weinrib, *Corrective Justice* (OUP 2012) 109.

the same injustice',[134] such that any permissible consideration must take account of one party's position *and* the other party's correlative position. In other words, any factor relevant to restoring the prior equilibrium must relate equally to the party who has lost, and the party who has gained, from the injustice. As the consideration must be relevant as between the parties, neither party can complain that his, her or its interests have been sacrificed to advance the other party's interests.[135] The result is that the corrective justice norm is fair to both parties.

The corollary of the fairness principle is that it is inappropriate for courts adjudicating private law claims to consider factors that are external to the parties' relationship. In particular, any consideration that is unilaterally advantageous, or disadvantageous, to either party is irrelevant in legal decision-making. Thus, the private law cannot determine a matter by considering solely the claimant's or defendant's side of the equation. Taking such a one-sided approach to determining liability would be incompatible with the notion of party equality embedded in the private law. As such, it would violate the fairness principle, and the corrective justice norm. For example, consider the factor of deterring wrongdoing. On this view, a court cannot rightly take into account this consideration because it focuses entirely on influencing the conduct of the defendant (or potential defendants), and hence it ignores the claimant. Consequently, basing liability upon a deterrent rationale is unfair.

In short, corrective justice theory disqualifies unilateral considerations referring to the position of only one party.

B Corrective justice and punishment

As illustrated below, corrective justice holds that punishment is unjustifiable in the private law of wrongs. Hence, it justifies the rule against punitive damages for a wrong and its underlying preventing punishment rationale. As previously explained, penalty clauses are a punitive damages analogue; in particular, both remedies function punitively. Given this link, when the private law prevents a party enforcing the former and when it does not award the latter, it acts based on a similar preventing punishment rationale. Accordingly, it seemingly follows that corrective justice norms justify both prohibitive principles and analogous rules. As we shall see, this is indeed the case.

Given the analogy between penalties and punitive damages, many corrective justice reasons against the private law recognising a punishment response as expressed through a right

134 Ernest J Weinrib, *Corrective Justice* (OUP 2012) 171.

135 Ernest J Weinrib, *Corrective Justice* (OUP 2012) 3.

to punitive damages also apply to justify why the private law does not recognise a punishment response as manifested in a right to a penalty. In short, many correctivist objections to wrongs-based punishment apply with equal force to consensual punishment.

This section contains two parts. It will first canvass the precedential landscape regarding the core categories of wrongs, and recapitulate on the law regarding consent. Then, the section will consider three principled corrective justice reasons. It will show how each reason supports the view that punishment generally is not, and in any event should not be, an available response to a wrong; and how each reason similarly supports the position that punishment is not, and should not be, a legitimate response to consent. Taken cumulatively, these correctivist reasons explain why punitive damages are unacceptable in the private law; and why punitive penalties are unenforceable.

I Precedent

(1) Wrongs

The boundaries of the law of wrongs are controversial. Suffice it to say that the category of wrongs uncontroversially includes breach of contract, equitable wrongs and torts. Each category is examined in order.

We begin with breach of contract. For over a century, Anglo-Australian private law has prohibited general punitive damages for breach of contract.[136] In *Johnson v Unisys Ltd*, Lord Steyn summarised the English law position thus: punitive damages 'have never [been] and cannot be awarded for breach of any contract'.[137] Equally, the Australian High Court has on several occasions confirmed that punitive damages are not to be awarded for a breach of contract.[138] Tettenborn concludes that the authorities deciding this issue are 'venerable and thus far unquestioned'.[139]

Turning to equitable wrongs, punishment is not a permissible response and hence punitive damages are unavailable. In *Harris*, the majority held that courts have no power to make a punitive award for a breach of fiduciary duty; and Heydon JA in obiter went further,

---

[136] *Addis v Gramophone* [1909] AC 488 (HL) 496.

[137] [2001] UKHL 13, [2003] 1 AC 518 [15].

[138] eg *Gray v Motor Accident Commission* [1998] HCA 70, (1998) 196 CLR 1 [13].

[139] Andrew Tettenborn (gen ed), *The Law of Damages* (2nd edn, LexisNexis 2010) [2.38].

holding that courts have no power to award punitive damages for any equitable wrong.[140] In addition, the English High Court and the Victorian Court of Appeal have each accepted that breach of confidence does not sound in punitive damages.[141]

Lastly, one must consider torts. Here, it is necessary to distinguish between the Australian and English positions.

In Australia, punitive damages are presently available for various torts, including trespass to the person, to land, and to goods.[142] However, it is important to note the Australian High Court's decision in *Gray v Motor Accident Commission*.[143] Although the joint judgment held that in principle punitive damages are available, it was at pains to stress that the Court was not asked to reopen the question of whether punitive damages properly should be available in tortious actions; or to consider whether the 'anomalies and difficulties that attend the awarding of [punitive] damages' in tort were 'such as to invite some … change to the law'.[144] Accordingly, these statements suggest that punitive damages may not be as entrenched in Australian private law as some might suppose. That said, it is undeniable that as the law presently stands punitive damages are awarded for some torts. To this, the appropriate reply is that Australian law on this issue is unprincipled.

In England, punitive damages are currently available for torts, but the circumstances in which a court will award them are severely restricted. English courts may award punitive damages for torts in only two situations: where a government servant has engaged in oppressive, arbitrary or unconstitutional wrongdoing; and where the defendant has cynically calculated that his or her conduct will make him a profit that may exceed the compensation payable to the claimant.[145]

In England, punitive damages are extremely controversial. Many judges and jurists take the view that punitive damages are anomalous, and should be expunged from the private law. That view is correct. The only justifiable position is not that punitive damages should be available, but awarded sparingly; rather, it is that punitive damages should not available at all.

---

[140] *Harris v Digital Pulse Pty Ltd* [2003] NSWCA 10, (2003) 56 NSWLR 298 [5], [44], [57], [233], [470].

[141] *Mosley v News Group Newspapers Ltd* [2008] EWHC 1777 (QB), [2008] EMLR 20 [186]-[197]; *Giller v Procopets* [2008] VSCA 236, (2008) 24 VR 1 [156]-[157].

[142] Keith Mason, JW Carter and Greg Tolhurst, *Mason and Carter's Restitution Law in Australia* (2nd edn, LexisNexis Butterworths 2008) [1510].

[143] [1998] HCA 70, (1998) 196 CLR 1.

[144] *Gray v Motor Accident Commission* [1998] HCA 70, (1998) 196 CLR 1 [10]. See also [33].

[145] *Rookes v Barnard* [1964] AC 1129 (HL) 1226.

Importantly, if English private law took this step, many decisions involving punitive damages could be saved, as they are capable of being explained on another basis. This applies to cases falling in each of the two categories.

So, many cases falling within the first category are better explained as involving a different type of damages, and hence a different response. For instance, consider *Benson v Frederick*.[146] A colonel ordered a common solider to be flogged so as to spite a fellow officer. The Court awarded the solider general damages of £150. In *Rookes v Barnard*, Lord Devlin cited the decision as a clear example of a punitive damages award in the first category.[147] However, the remedy is better understood as an aggravated damages award that compensated for the injury to the soldier's moral dignity, on the basis that 'though not much hurt indeed, [he] was scandalized and disgraced' by the flogging.[148]

In addition, regarding several cases falling within the second category, the remedy granted is better interpreted as a disgorgement damages award, implementing a gain-based response. Several judges have acknowledged this. For example, in *McCarey v Associated Newspapers Ltd*, Diplock LJ held that if a court makes an award pursuant to this category, 'This is not punishment; it is merely preventing the defendant from obtaining a reward for his wrongdoing.'[149] Further, in *Kuddus v Chief Constable of Leicestershire Constabulary*, two members of the House of Lords held that this category was unpersuasive, and that in any event it had been 'largely overtaken' such that disgorgement damages were the appropriate relief in the instant circumstances.[150] Moreover, in *Kuddus*, Lord Nicholls noted that it is anomalous that a profit motive suffices to ground a punitive damages award, but a malicious motive does not.[151] The concern with a defendant's profit motive suggests that the law is, in truth, concerned with preventing illegitimate profit making. If that is correct, then a profit stripping remedy more fully achieves this aim.

In conclusion, punitive damages are unavailable for the wrong of breach of contract and equitable wrongs, and but are (restrictively) available for tortious wrongs. To be coherent, the Anglo-Australian private law of wrongs must take a consistent approach to punitive

---

146 (1766) 3 Burr 1845, 97 ER 1130. See also *Huckle v Money* (1763) 2 Wils KB 205, 95 ER 768 (a case of substitutive damages for the value of the infringed right to freedom of movement).

147 [1964] AC 1129 (HL) 1222-23.

148 *Benson v Frederick* (1766) 3 Burr 1845, 1846; 97 ER 1130, 1130.

149 [1965] 2 QB 86 (CA) 107.

150 [2001] UKHL 29, [2002] 2 AC 122 [67], [109].

151 *Kuddus v Chief Constable of Leicestershire Constabulary* [2001] UKHL 29, [2002] 2 AC 122 [67].

damages. If one considers the precedents regarding punishment for wrongs as a holistic category, it is the law of torts that jars with the rest of the law in this class. Accordingly, the best way to render the Anglo-Australian private law of wrongs coherent is to ensure that punishment as expressed though punitive damages is an impermissible response throughout the entire category. For breach of contract and equitable wrongs, this only requires a continued acceptance of what the law is; for torts, it requires a move to what the law should be. In England at least, the impact of such a change would not be dramatic.

(2) Consent

Chapters 3 to 6 addressed the question of whether, as a matter of doctrine, consent can generate a right to a punitive remedy. It was shown that the Penalty Rule has existed for centuries. It continues to thrive today. Accordingly, the answer to the question posed is an emphatic 'no'.

II Principle

(1) Preliminaries

Before examining the principled arguments, we must note three fundamental points.

First, on the orthodox view, a defendant's obligation to pay general damages for a breach of a primary duty is a manifestation of corrective justice. Indeed, the corrective justice explanation of secondary rights has been labelled 'trite', because it is self-evident that all rights infringements require correction.[152] Similarly, on the view presented here, B's obligation to pay A a liquidated damages sum for the breach of a primary obligation is itself a primary obligation, and also a manifestation of corrective justice.

So, recall that a right to liquidated damages arises from a consent-based actual wrong. Thus, where the event triggering B's duty under a liquidated damages clause is a breach of contract, B's duty is a primary duty arising from consent that exhibits the same structure or anatomy as a general law secondary duty arising from a wrong. Applying a formalist analysis,[153] the private law will treat B's primary duty in the same way as a defendant's general law secondary duty, because B's primary duty exhibits the same genetic make-up as a general law secondary duty.

---

[152] Robert Stevens, *Torts and Rights* (OUP 2007) 327.

[153] Text following n 37 in ch 1.

Viewed through a corrective justice lens, the analogical link between a right to general damages and a right to liquidated damages becomes apparent. When a defendant commits a general law wrong, such as a breach of contract, he violates the corrective justice norm and it creates an injustice. To correct that injustice, the private law responds with a general law secondary duty. Equally, when B commits a consent-based actual wrong, he violates that same corrective justice norm and it generates an injustice. To correct that injustice, the parties respond with a consensual primary duty, which obligation takes the same shape as a general law secondary duty. The two duties display the same form because they share the same normative foundation. It is incorrect to assert, then, that the Penalty Rule rests on the view that secondary duties are the 'preserve of the courts'.[154] Rather, all secondary obligations and primary obligations of a like form are the preserve *of corrective justice*.

Second, Smith has argued that neither the judges administering the law, nor the parties, can select the consequences of a violation of corrective justice; the consequences must satisfy the requirements of corrective justice itself.[155] Put differently, only corrective justice can determine the acceptable response to a violation of its norms. This has implications for the law of wrongs, and the law of consent.

As to the law of wrongs, corrective justice constrains the permissible responses to any breach of obligation in the private law. Thus, corrective justice holds a monopoly over the response to a wrong: compliance with its precepts is *mandatory*. For instance, as Collins has written, the traditional understanding of general damages for breach of contract is that they 'must be confined to corrective justice'.[156] Consequently, judges administering the law are not free to depart from what corrective justice requires.

As to the law of consent, a consent-based breach violates the corrective justice norm. Given this, only corrective justice can tell us what constitutes an acceptable response to that violation. Thus, as is the case with general law wrongs, corrective justice holds a monopoly over the permissible responses to a consent-based wrong. Accordingly, when invoked, corrective justice precepts represent mandatory norms that bind the parties. Hence, Smith argues that the Penalty Rule implements 'a mandatory [norm] which denies the parties full autonomy in crafting the consequences of a breach of primary obligations'.[157] Similarly,

---

[154] Mindy Chen-Wishart, 'Controlling the Power to Agree Damages' in Peter Birks (ed), *Wrongs and Remedies in the Twenty-First Century* (Clarendon Press 1996) 273.

[155] Lionel Smith, 'Relief Against Forfeiture: a Restatement' (2001) 60 CLJ 178, 188.

[156] Hugh Collins, *The Law of Contract* (4th edn, LexisNexis 2003) 422.

[157] Lionel Smith, 'Relief Against Forfeiture: a Restatement' (2001) 60 CLJ 178, 185 fn 18.

Worthington has noted that the values the rule protects 'are currently seen as overriding freedom of contract'.[158]

Finally, as the corrective justice norm is mandatory, any purported response to a wrong that fails to comply with that norm will be prohibited. A judge simply will not award a remedy that instantiates such a response. As the following illustrates, this is why properly understood there can be no punishment for a general law wrong. Equally, the mandatory nature of the corrective justice norm means that the parties must comply with its dictates. If they fail to do so, the court will not enforce the right that the parties have sought to create. The parties cannot, under any circumstances, 'contract out' of what corrective justice demands.[159] As the following will show, this explains why the private law does not sanction a right to consensual punishment.

(2) Fairness

(a) Wrongs

An award of punitive damages for a wrong violates the fairness principle because it ignores the claimant.

To explain, when a court awards punitive damages its focus is not on the relationship between the defendant's act and the claimant's right, but 'on the defendant's conduct alone. The restoration of what is rightfully the [claimant's] is therefore not a reason for awarding the [claimant] this variety of damages.'[160] In other words, punishment fixates on the defendant and completely neglects the claimant. Punishment focuses solely on the defendant's behaviour, and in particular on ensuring that the defendant is deprived of something due to his or her wrongdoing. Accordingly, it is not a liability consideration that relates to the parties' relationship *as a whole*, for it cares nothing for the claimant. It constitutes a one-sided normative consideration. The criminal law proves this. In some cases, the criminal law punishes failed attempts to commit offences. Courts are only able to do this because punishment focuses exclusively on the wrongdoer's misconduct, rendering the existence of a victim irrelevant where punishment is otherwise warranted.

---

158 Sarah Worthington, 'What is Left of Equity's Relief Against Forfeiture?' in Elise Bant and Matthew Harding (eds), *Exploring Private Law* (CUP 2010) 272 (footnote excluded).

159 Lionel Smith, 'Relief Against Forfeiture: a Restatement' (2001) 60 CLJ 178, 189.

160 Ernest J Weinrib, 'Two Conceptions of Remedies' in Charles Rickett (ed), *Justifying Private Law Remedies* (Hart Publishing 2008) 24.

Punishment is, therefore, unilateral: it 'focuses not relationally on the parties as doer and sufferer of the same injustice, but unilaterally on the defendant as doer'.[161] Thus, a judicial remedy that encapsulates a punitive response is unfair as between the parities, and cannot stand with the fairness principle.

(b) Consent

Equally, a right to a penalty transgresses the fairness principle, as in all cases it ignores A.

A penalty clause operates punitively, by punishing B for his default. Any punishment for a consent-based wrong is necessarily a one-sided normative consideration, as it exclusively focuses on B's conduct whilst paying no regard to A. This remains true even if punishment is used as a means to the ends of deterring B from committing that wrong, and even if A has agreed to the clause—A's agreement cannot change the nature of the concept of punishment (or deterrence), which is inherently unilateral.

In short, a penalty clause is impermissible because it is distinctively one-sided in its focus on B's behaviour. Given its punitive nature, the clause fixates upon B, and wholly neglects A. Consequently, the right that arises from such a clause is unfair as between the parties, and transgresses the fairness principle.

(3) Correction

(a) Wrongs

A punitive response to a wrong violates the correction principle, as it does not accord with the equivalency thesis or the limitation thesis.

As to the former, according to the corrective justice norm, the just response to a wrong is to place the claimant into the next best position. An award of punitive damages goes beyond this next best position. So, an award of several other forms of general damages notionally will place the claimant in the required counterfactual position. For instance, where a defendant has breached his contractual obligation to the claimant and caused her to suffer consequential loss, a general compensatory damages award will place the claimant in the position she would have occupied if the defendant had performed his obligations, in that she will be placed in the same factual position. However, a punitive damages award is necessarily

[161] Ernest J Weinrib, *Corrective Justice* (OUP 2012) 171.

parasitic upon, and in excess of, that other general damages award. As such, notionally it will leave the claimant in a better position than if she had not suffered any wrong at all. Thus, in the example, an additional punitive damages award will leave the claimant's balance sheet swollen as compared with that which would have existed if the defendant had performed. In that situation, the claimant obtains a windfall. Importantly, it is a fine extracted *from the defendant* that funds this windfall.[162]

As to the latter, Weinrib argues that the limitation thesis requires that a remedy should 'only restore the [claimant's] right and not give the [claimant] more than that right (or its equivalent)', and that accordingly punitive damages transgress this thesis.[163] Many kinds of general damages awards restore to the claimant in full the monetary equivalent of her primary right. However, the corrective justice norm prohibits the private law going further. This is precisely what a punitive damages award seeks to do. By their very structure, punitive damages are excessive; they go 'above and beyond' the money sum that represents the primary right in a financial form. Consequently, punitive damages surpass the normative limits which corrective justice sets, and which properly underpin any justifiable general damages remedy.

(b) Consent

Similarly, a consensually punitive response transgresses the correction principle, as it does not adhere to the equivalency thesis or the limitation thesis.

As to the equivalency thesis, a fundamental feature of that thesis holds that a secondary right to general damages instates the next best position. As the agreed damages regime is an analogue of the general damages regime, the parties consensual right to liquidated damages—which is primary in source but secondary in form—must in some sense *reflect* the next best position. As Burrows acknowledges, 'courts are not hostile to liquidated damages because they reflect the … aim of judicially assessed damages: that is, they are a genuine pre-estimate of the sum needed to put the claimant into as good a position as if the contract had been performed'.[164] Liquidated damages are a sum representing an *estimate* of the amount likely needed to put A into as good a position as if B had performed his obligation; that is, an estimate of the amount probably needed to put A into the next best position. Therefore, liquidated damages constitute a sum seeking to place A into the 'next-next best' position. The

---

[162] *Cassell and Co Ltd v Broome* [1972] AC 1027 (HL) 1086. See similarly *Harris v Digital Pulse Pty Ltd* [2003] NSWCA 10, (2003) 56 NSWLR 298 [345].

[163] Ernest J Weinrib, *Corrective Justice* (OUP 2012) 97.

[164] Andrew Burrows, *Remedies for Torts and Breach of Contract* (3rd edn, OUP 2004) 449.

grant of a punitive sum in the form of a right to a penalty, however, goes beyond the next-next best position.

The next-next best position constitutes an appropriate approximation of the amount likely needed to put A into the next best position. It is only acceptable to place a party into the next-next best position if the next best position is unknowable. This is the case with liquidated damages, as they are quantified *on the contract date* and are therefore necessarily prospective in nature.

So, suppose we lived in a world where a crystal ball was fact, and not fiction. The judge would be able to say that, on the contract date, both parties would have known beyond peradventure what the consequences of B's breach would be. If so, the private law could demand that the agreed damages sum be precisely the same as the general damages sum. Put differently, the Penalty Rule might require that the parties instate the next best position. However, if we accept that the parties can choose to have matters determined *ex ante*, and that we live in a world without magic, then on the contract date the next best position is necessarily unknowable. Accordingly, the best the court can do is require that the parties select an amount that is the closest to the next best position that one can *realistically* achieve, given that future events are difficult to predict. In this situation, the only *workable* standard is one such as 'extravagance', which encapsulates a wide margin of appreciation. In the circumstances, the closest one can get to the best position is the next-next best position, for that is the only realistic and workable option. Conversely, it would be entirely unrealistic to require the parties to instate the *next best* position (for example, by requiring that the agreed damages sum represent a substantially accurate pre-estimate of loss). An unrealistic option cannot represent the closest one can get to the best position, for an unrealistic option is not an option at all.

Thus, the best the private law justly can do is this: it can require that an agreed damages clause compels B to pay a sum which places A in a position which is as close as possible to the position of the wrong not having occurred, *given the anticipatory nature of the exercise*. This means that the parties must install the next-next best position, which comprises an acceptable approximation of the amount probably needed to put A into the next best position. In many cases, this approximation may not be very close at all to the replicating the next best position. But, given the assessment date, it is the best realistically that can be done.

It follows that a penalty clause is unenforceable because it goes beyond the next-next best position. So, the punitive component of a penalty sum is necessarily parasitic upon, and in excess of, the component of that sum that implements another response, such as substitution, compensation or disgorgement. Given this, notionally the penalty sum will leave A better off: it will provide A with an amount exceeding any amount that, assessed at the contract date, she will likely need to place her into the next best position. In that sense, A obtains a windfall; and B is forced to pay a fine to fund that windfall.

As to the limitation thesis, any right to liquidated damages must, in an appropriately analogous way, represent the original primary right viewed 'from the other end'. In other words, it must constitute the monetary equivalent of the original primary right, assessed prospectively. Corrective justice prohibits the parties from granting to A a right to recover any more than this sum. This, however, is exactly what a right to a penalty attempts to do. By definition, a penal sum contains an excessive element that seeks to go 'above and beyond' the amount that reflects the original primary right in money terms. By reason of that component, penalty clauses contravene the limitation thesis.

(4) Monism

(a) Wrongs

A secondary right to punitive damages exists outside the structure of civil liability, and thus violates the monism principle.

So, according to Beever, a fundamental objection to punitive damages is that they operate punitively, and that 'punishment is foreign to the structure of private law'.[165] The private law's structure is said to leave no room for punishment. Accordingly, the writer argues that it is impossible to integrate punitive damages into the framework of civil liability in an intelligible and coherent fashion. Punitive damages, then, are a fish that walks the private law's terrain. They do not make sense. To make good this claim, three steps are required.

To start, it is necessary to identify the correct structure of a private law claim for a wrong. Correctivists articulate that structure as follows.[166] The defendant owes the claimant a primary duty. If the defendant breaches that duty, it engenders a legitimate response, which is typically compensation. The secondary duty raised in the defendant that implements this response 'fits' the primary duty previously owed. Put differently, it is the duty breached that defines the remedy. Importantly, the claimant (and she alone) holds the secondary claim right

---

[165] Allan Beever, 'The Structure of Aggravated and Exemplary Damages' (2003) 23 OJLS 87, 105.

[166] Allan Beever, 'The Structure of Aggravated and Exemplary Damages' (2003) 23 OJLS 87, 106; CEF Rickett, 'Punitive Damages: the Pulse of Equity' (2003) 77 ALJ 496, 498.

that correlates to this secondary duty. She may only exercise this right against the defendant (and he alone). In that way, the private law operates as between the parties.

The reasoning here neatly reflects the monism principle. On the monist view, the secondary right is the primary right admired 'from the other end': it is a normative continuation of the primary right. This is why, although the secondary right takes a different form, it nevertheless 'fits' the anterior primary right.

Next, after the private law has awarded a secondary right to general damages implementing a relevant response—such that the claimant's loss is repaired, for instance—the injustice is undone, and the ledger is wiped clean. Thus, there is no residue. As such, the private law 'runs out'. It follows that if the defendant properly owes a secondary duty to pay a punitive sum, that duty cannot be a response to the defendant's breach of a primary duty *owed to the claimant.* Accordingly, a punitive damages award cannot be justified as part of the legal relations between the parties. Such an award must respond to the defendant's breach of a duty of a different kind.

Lastly, given this reasoning, Beever argues that the secondary duty to pay punitive damages arises from a defendant's breach of a primary duty owed to *society at large*, not one owed to the individual claimant. According to the writer, society will not tolerate an individual breaching his or her primary duty owed to another individual in a serious way. As a result, in punitive damages cases, the defendant has infringed a primary right of the public[167]—not the claimant—to have an individual refrain from seriously breaching his or her duty owed to another individual. The defendant's secondary duty to pay punitive damages, then, is a duty owed to the general public. It is not a duty owed to the claimant. In Beever's words, 'the duty is owed to society at large. Perhaps it is owed *when* defendants seriously breach duties owed to claimants, but it is not owed *for* those duties. Liability for [punitive] damages … is not a "term of relation" between the parties, it results from "a wrong to the public at large".'[168]

Importantly, a fundamental element of punitive damages suggests that they are irrelevant to the structure of private law wrongs claims. As Chapter 1 noted, a critical feature of civil punishment, as instantiated in punitive damages, is that it is always dispensed for *the way* in which the defendant breached a particular duty, rather than the breach itself. A court will only award punitive damages if the *quality* of a defendant's conduct when committing the

[167] 3 Bl Comm 2 cf Robert Stevens, 'Private Rights and Public Wrongs' in Matthew Dyson (ed), *Unravelling Tort and Crime* (CUP 2014) 113.

[168] Allan Beever, 'The Structure of Aggravated and Exemplary Damages' (2003) 23 OJLS 87, 107 (footnotes omitted; original emphasis).

breach is sufficiently serious, in that it displays an outrageous and contumelious disregard for the claimant's rights.[169] However, the way in which a defendant breaches a relational duty, or the quality of his conduct in so doing, is properly irrelevant to the structure of private law liability. The private law is only concerned with the rights infringement *per se*, and its loss-based or gain-based consequences. The manner in which the right is infringed is not relevant to any interpersonal injustice, or its correction.

Due to the monistic nature of the relationship between primary and secondary rights, a breach of duty owed to another individual can never generate a punitive response. Accordingly, on a correctivist view, punitive damages are 'incapable of being coherently integrated into the fabric of [private law] liability'.[170] A civil fine is unjustifiable. Only the criminal law may justifiably dispense fines, and then only for a wrongdoer's breach of duty owed to society at large. Undoubtedly, sometimes the nature of the defendant's behaviour will call for a punitive response from the law. But that is a dance that the private law must sit out.

(b) Consent

Likewise, a right to a penalty stands outside the structure of civil liability, and hence transgresses the monism principle.

As discussed, a defendant's secondary duty to pay punitive damages is owed for committing a wrong to society at large, such that punitive damages are a civil fine. Does the punitive element of a penal sum similarly represent a purported sanction for the breach of a duty owed to the public in general? It does. So, Rossiter argues that a good justification for the Penalty Rule lies in the proper divide between the role of the private and public law: 'It is the function of the state, through its criminal laws, to punish and fine; this role must not be delegated to … citizens in their private capacity …'.[171] Similarly, Rickett has contended that the rule is concerned to counter an unacceptable insistence on the exercise of rights because 'the pursuit of punishment is a matter for the state and not private law …'.[172]

In essence, a penalty clause is impermissible because it partly seeks to provide a response that is reserved to the public law, to be dispensed only where a wrongdoer has infringed a public right. Accordingly, the reason why all forms of civil punishment (such as

---

169 *Kuddus v Chief Constable of Leicestershire Constabulary* [2001] UKHL 29, [2002] 2 AC 122 [63].

170 Ernest J Weinrib, *Corrective Justice* (OUP 2012) 183.

171 Chris Rossiter, *Penalties and Forfeiture* (Lawbook Company 1992) 33.

172 Charles Rickett, 'Unconscionability in Commercial Law' in John Lowry and Loukas Mistelis (eds), *Commercial Law: Perspectives and Practice* (Lexis Nexis Butterworths 2006) [9.27].

punitive damages and penalties) are prohibited is that properly they can relate only to a wrong done to society, not to an individual. Therefore, an individual has no place punishing another individual, by extracting a fine or otherwise. As it turns out, then, 'civil punishment' is in truth an internally contradictory concept.

There is one additional point. As previously noted, a central feature of civil punishment is that it is always dispensed for the way in which an individual breached a particular duty. As illustrated in the punitive damages context, the quality of an individual's conduct when committing a breach is immaterial to the structure of private law liability. Equally, given the agreed damages regime is an analogue of the general damages regime, the way in which B breached his original primary duty is properly irrelevant to the structure of consent-based private law claims.

C The a fortiori case

Several corrective justice reasons rendering punishment imposed for a wrong unacceptable similarly pertain to punishment agreed by consent. However, a further reason exists that exclusively applies to consensual punishment to undermine its normative acceptability. Hence, in terms of justifying the underlying prohibitive principle, consensual punishment is a fortiori wrongs-based punishment.

Importantly, punishment is a type of response, but it does not *of itself* provide a justification for that response. Two traditional justifications of punishment exist: deterrence and retribution.[173] Each is arguably a manifestation of distributive justice.[174] Obviously, neither provides a valid justification from the corrective justice standpoint.

Assume that the principle against punishment for a wrong did not exist, and that punitive damages were available for all wrongs. If this occurred, courts would most likely justify punishment as expressed through a right to punitive damages based on deterrence.[175] In such a world, although deterrence would per hypothesis justify punitive damages, it could not

---

[173] *Uren v John Fairfax and Sons Pty Ltd* (1966) 117 CLR 118 (HCA) 149. See also *Wilkes v Wood* (1763) Lofft 1, 18-19; 98 ER 489, 498-99 (deterrence); *A-G of Trinidad and Tobago v Ramanoop* [2005] UKPC 15, [2006] 1 AC 328 [19] (retribution).

[174] On one view, retribution is ultimately part of distributive justice: Allan Beever, 'Justice and Punishment in Tort: A Comparative Theoretical Analysis' in Charles Rickett (ed), *Justifying Private Law Remedies* (Hart Publishing 2008) 269-70.

[175] James Edelman, 'In Defence of Exemplary Damages' in Charles Rickett (ed), *Justifying Private Law Remedies* (Hart Publishing 2008).

similarly justify a penalty clause. This is because, unlike wrongs-based punishment, consensual punishment will have no real generally deterrent effect.

To explain, Edelman argues that deterrence is the correct justificatory ground for imposing punishment as expressed through punitive damages. But the relevant deterrence is of a particular brand. Thus, in *Lamb v Cotogno*, the Australian High Court awarded punitive damages despite the fact the defendant's insurer was to make the payment.[176] The fact that the defendant's insurer—and not the defendant—would pay the punitive sum necessarily meant that the award would not specifically deter the defendant. However, it was said that other 'like-minded' individuals willing to 'engage in conduct of a similar nature' might be uninsured.[177] Accordingly, Edelman argues that *Lamb* provides an example of a court awarding punitive damages 'solely to deter others'.[178] Therefore, in the writer's view, it is *general* deterrence that justifies a right to punitive damages. On a deterrence-based view, then, punitive damages can create a generally deterrent effect, which is sufficient to justify the award.

Contrast a penal sum agreed by consent. At best, general deterrence is only minimally relevant. At worst—and the most likely case—general deterrence is irrelevant. It is difficult to see how upholding an *inter partes* remedy, contained in commercial agreement, will deter the conduct of similarly situated parties in future. Two reasons follow.

For one, assume that courts upheld penalty clauses, such that litigation to enforce these clauses was unnecessary. Given the private nature of commercial agreements, in this situation the general commercial public would be unlikely to become aware of many, if not all, consensually punitive sanctions. As most members of that group would be left unaware, they could not be deterred from acting similarly. Any sanction will not deter people generally if the facts of the case are not widely known.

For another, consensual punishment does not carry the same stigma as imposed punishment. Accordingly, its generally deterrent effect is marginalised. So, suppose that X and Y entered into a contract containing a penalty clause. X then breaches his contract with Y, and pays a large penalty amount. One year later, A and B are negotiating a nearly identical contract. Given a consensual punishment does not carry as much obloquy as an imposed punishment, B may discount the fact that X breached his obligation to Y, and subsequently paid a large punitive sum. Consequently, B may be less discouraged from breaching his duty

---

176 (1987) 164 CLR 1 (HCA) 9.

177 *Lamb v Cotogno* (1987) 164 CLR 1 (HCA) 10.

178 James Edelman, 'A "Fusion Fallacy" Fallacy?' (2003) 119 LQR 375, 377.

than if a court had made an order against X in the same circumstances. This point holds good even if, contrary to current practice, the commercial community began to publicise widely the details of private agreements containing consensual punitive sanctions.

If the private law awarded punitive damages for wrongs, the law's scope to 'punish *pour décourager les autres* ['to discourage the others'] in significant numbers' would either justify that punishment, or at least strengthen the case for it.[179] Conversely, general deterrence can provide only a very weak justification for private punitive practices, if any at all. Realistically, a penalty clause will not produce a generally deterrent effect. Rather, a right to a penalty will create mostly—or only—a specifically deterrent effect.

The short point is this. The right to punitive damages and the right to a penalty are analogous, proscribed rights. However, the prohibition against the latter is stronger. As far as the preventing punishment principle goes, consensual punishment is the a fortiori case.

3 Conclusion

This chapter illustrated that the Penalty Rule's proper underlying principle is one of preventing consensual punishment; and that this principle instantiates a more profound normative rationale of satisfying corrective justice. Accordingly, corrective justice provides the rule's ultimate theoretical explanation. Thus, a classificatory, corrective justice theory provides an intelligible and powerful reason for the Penalty Rule's existence.

---

179 John Gardner, 'Introduction' in HLA Hart, *Punishment and Responsibility: Essays in the Philosophy of Law* (John Gardner ed, 2nd edn, OUP 2008) l.

9 EVALUATION

This thesis has propounded a two-level theory of the Penalty Rule. At the low level, the classificatory theory of the rule holds that it embodies a principle against consensual punishment. At the high level, the corrective justice theory of the doctrine maintains that correctivist norms justify that principle. Accordingly, the Penalty Rule ultimately instantiates corrective justice.

This chapter critically assesses that dual theory. It first adopts the three interpretive axes of fit, transparency and coherence to evaluate the theory. The chapter then considers the ramifications of practitioners embracing this rights-based thesis.

This chapter establishes that the propounded rationale performs better on each criterion than the other putative justifications for the rule. It follows that this understanding of the rule best explains it, and powerfully justifies it.

1 Fit

A classificatory, correctivist explanation of the Penalty Rule generates outcomes that are consistent with the decisions in the cases.

The two-level theory of the rule completely satisfies fit in relation to the 'central' cases. For example, on this theory, an orthodox compensatory analysis produces the outcomes in the 'central' cases of *AMEV-UDC Finance Ltd v Austin*[1] and *Makdessi v Cavendish Square Holdings BV*;[2] and a substitutive approach achieves the result in the 'central' cases of *Clydebank Engineering and Shipbuilding Co Ltd v Don Jose Ramos Yzquierdo y Castaneda*[3] and *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd*.[4] In addition, this dual theory satisfies fit with respect to nearly all of the 'standard' and 'controversial' cases. For instance, a substitutive analysis generates the outcome in the 'standard' case of *State of Tasmania v Leighton Contractors Pty Ltd*,[5] and a gain-based approach leads to the decision in the 'controversial' case of *R v A-G for England and Wales*.[6] Notably, given this theory is rights-based, it is consistent with the

---

[1] (1986) 162 CLR 170 (HCA).

[2] [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125.

[3] [1905] AC 6 (HL).

[4] [1915] AC 79 (HL).

[5] [2005] TASSC 133, (2005) 15 Tas R 243.

outcome of the 'controversial' case of *Cellulose Acetate Silk Co Ltd v Widnes Foundry (1925) Ltd.*[7] In short, this theory displays a very tight fit.

Importantly, the Penalty Rule has several fundamental features that the leading theories perceive to be anomalous, or fail to explain at all. Conversely, the classificatory, correctivist theory can rationalise these features. Three examples will suffice.

A The breach requirement

As Chapter 7 showed, for the substantive unconscionability theory the Penalty Rule jurisdictional test is a mystery. It cannot explain why the breach requirement exists; why the paradigm scenario[8] involves a breach; or why a substantive breach is properly sufficient to engage the rule.

Conversely, a corrective justice theory can explain the jurisdictional rule's distinctive features. The following considers first the 'core case' of a consent-based actual wrong, and then a consent-based substantive wrong.

I Consent-based actual wrongs

Corrective justice theory explains the breach requirement on the following basis: a breach of contractual duty violates correctivist norms by disrupting the parties' pre-existing equilibrium, and hence creating an injustice requiring correction.

To explain, Chapter 8 argued that when B commits a consent-based actual wrong against A, he violates corrective justice norms, such that an injustice arises. To correct that injustice, the parties respond by imposing a consensual primary duty upon B, which obligation takes the same shape as a general law secondary duty arising in response to a general law wrong. The two duties have the same genetic make-up because they respond to the same type of occurrence, namely an injustice; and because they share the same normative basis, as they both instantiate the response which corrective justice demands. This explains why the general law demands a breach, and why the Penalty Rule review jurisdiction contains a breach requirement: a wrong is fundamental because it provides the 'core case' of an injustice, and an injustice is a fulcrum of corrective justice theory. Without it, there is simply nothing to correct.

The relevant duty that arises to correct the injustice is a *primary* duty, not a secondary

[6] [2002] 2 NZLR 91 (NZCA), affd [2003] UKPC 22, [2003] EMLR 24.

[7] [1933] AC 20 (HL).

[8] Text following n 20 in ch 1.

duty. For two reasons, this is unproblematic.

For one, the conventional view is that the corrective justice norm applies to some primary duties (such as a duty to make restitution in unjust enrichment).[9] What matters is not whether the duty is primary or secondary, but 'the structure of the injustice and the consequent structure of the remedy'.[10] In short, it is the nature of the event, and whether it displays the structure of an 'injustice', that determines whether a response must satisfy corrective justice.

For another, although the duty here is primary, it is normatively distinct because of its form, as it exhibits the same structure or anatomy as a general law secondary duty. Duties of this shape lie at heart of corrective justice theory. If general law secondary rights manifest corrective justice, so too must consensual primary rights that bear the same structure as those rights.

II Consent-based substantive wrongs

Corrective justice theory rationalises why the jurisdictional rule catches substantive breaches, on the basis that the parties cannot stultify mandatory corrective justice norms. A little unpacking is necessary.

In Chapter 4, this thesis noted that the review power extends to some clauses that strictly do not involve a rights infringement. Further, it contended that a concern for anti-stultification properly justifies this stance. That is, if A adopts an avoidance technique to circumvent the jurisdiction, failing to review the clause would make nonsense of the private law's commitment to reviewing agreed damages clauses with a view to invalidating penalty clauses. Consequently, the Penalty Rule would be stultified. Conversely, preventing stultification gives full effect to the breach requirement, ensuring that courts review all clauses that truthfully can be said to involve a consent-based breach.

How does the anti-stultification rationale fit with the corrective justice theory of the Penalty Rule? The answer is this. Smith contends that the parties cannot, under any circumstances, 'contract out' of corrective justice.[11] It is a mandatory norm that *always* decides how a legal injustice is to be rectified. However, if this mandatory norm is to have purchase, corrective justice *must* frustrate any attempts to circumvent it. Otherwise, this 'mandatory' norm is an empty gesture. Accordingly, corrective justice repels all stultifications, being

[9] Text accompanying and following n 119 in ch 8.

[10] Ernest J Weinrib, *Corrective Justice* (OUP 2012) 94.

[11] Lionel Smith, 'Relief Against Forfeiture: a Restatement' (2001) 60 CLJ 178, 189.

circumventions without a sufficient reason: in this case, a circumvention without sufficient reason of corrective justice's refusal to countenance punitive sanctions.

Thus, where the parties use technical drafting techniques to attempt to draft around the rule's proper jurisdiction, the private law enunciates what corrective justice demands. The law protects its monopoly over correcting injustices, and treats the clause based on what it truly involves: a consensual obligation arising upon an event that amounts to an inconsistency with A's right for which B is responsible. More concretely, where the parties have avoided creating a triggering event that takes the form of a consent-based actual wrong, the jurisdiction will catch the clause if that engaging event is, *in substance*, a consent-based actual wrong.

In conclusion, corrective justice always dictates the requirements of a response to a wrong *per se*; to a consent-based event that 'picks up' on the structure of a wrong; and to a consent-based event which purports to *avoid* 'picking up' on the structure of a wrong, thereby stultifying the demands of corrective justice. In this way, corrective justice provides a theoretical basis that justifies the scope of the jurisdictional test. It rationalises why the 'core case' involves a rights violation. Further, corrective justice explains why the jurisdictional scope, as properly understood, involves a substantive rights infringement.

B The assessment date

The Penalty Rule test of validity adopts a prospective assessment date. As Chapter 7 noted, the substantive unconscionability theory purported to explain why the rule adopts an *ex ante* assessment date, but failed. The other theories examined did not even make an attempt.

Lanyon argues that any theory that supports a preventing punishment principle necessarily cannot explain why the test of validity applies at the contract date.[12] That is wrong. The reason for why the test is, and must be, applied *ex ante* is that the parties impliedly agree to this assessment date; and they do so because they want the effect of certainty that comes from applying the test as at the contract date. An explanation follows.

I The effect of certainty

Like deterrence, certainty is not a *purpose* of the private law. Rather, it is an *effect* of the private law that comes into play after its norms have been determined.[13] So, in this context, the certainty in play is an effect of a liquidated damages clause being assessed, and validated, at the

[12] Elisabeth Lanyon, 'Equity and the Doctrine of Penalties' (1996) 9 JCL 234, 239.

[13] Ernest J Weinrib, 'Deterrence and Corrective Justice' (2002) 50 UCLA L Rev 621, 624-25, 633, 638-39.

contract date. As Chapter 6 noted, if the *purpose* of all agreed damages clauses was to create certainty, there would be no Penalty Rule at all: the law would simply enforce all agreed damages clauses out-and-out. Therefore, although certainty is a useful *consequence* of the test of validity operating in the way it does, it is not a *reason* for it operating in the way it does. Although an area of the private law might produce a number of incidental benefits, we should not confuse those benefits with what that area of law is 'about'.

II Implied agreement

It is by the parties' implied agreement that the effect of certainty becomes relevant. That agreement also leads to an *ex ante* assessment date.

So, as Chapter 6 explained, when the parties agree to an agreed damages clause, they do so seeking to obtain the certainty that a valid agreed damages clause will provide. The parties wish to be reasonably certain of the extent of each party's rights and obligations. If a dispute ensues, the only way of achieving this desired effect of certainty is to have the court evaluate the clause's validity as at the contract date, as it will mean that a court likely will uphold the clause.[14] Conversely, if a court were to assess the clause as at the breach date or judgment date, in a world without magic it would probably invalidate many commercial agreed damages clauses. As to liability, then, the strong likelihood of a certain effect would be replaced by the near certainty of an unpredictable effect. Accordingly, the parties' desire for a certain effect grounds their choice of this particular assessment date. It is implied in every case.

However, Anglo-Australian private law only applies the test of validity on the contract date because this is consistent with, and justified by, corrective justice. So, all general damages claims must comply with corrective justice. An agreed damages regime is merely an analogue of a general damages regime. It follows that an agreed damages claim must similarly comply with corrective justice norms. In the agreed damages context, the parties agree to apply those norms as at the contract date. They agree to do so because they want the effect of certainty that comes from this. That agreement is always implied. But despite the earlier assessment date, the norms applied *remain those of corrective justice*. The same principles will apply (such as those of injustice, correction and monism). In some cases, the same principle will apply in an identical fashion; in other cases, it will apply in an analogous way because of the changed circumstances.

[14] *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [75(iii)].

This is all possible because, for corrective justice, the temporal relationship between the injustice and its correction is irrelevant. That is, permitting the parties to select a particular assessment date maximises their personal autonomy and freedom of choice, but does no violence to the corrective justice norm, as the logic of that norm does not demand that its constituent principles must be applied at the breach date, or a later date. As Weinrib explains, corrective justice applies even when there is 'yet no wrong to correct', for what ultimately matters 'is not the temporal relation between the injustice and the remedy, but the structure of the injustice and the consequent structure of the remedy'.[15] Similarly, Gardner argues that although corrective justice binds one party to do the next best thing in any event, this does not mean the parties cannot agree upon the date that the next best thing will be assessed.[16]

In short, whenever the private law responds to a legal injustice, it must ask the question: 'What does corrective justice require?' It must ask the corrective justice question. A court will usually ask this question on the breach date, or judgment date. The way the test of validity works merely illustrates that the parties are permitted to have the court ask that corrective justice question as at the contract date. The time of asking may differ, but the question remains the same. In either case, the answer will meet the demands of corrective justice.

C The margin of appreciation

Chapter 5 contended that, properly understood, the Extravagance Test is the sole Penalty Rule test of validity. It further noted that the Extravagance Test inherently sanctions a large pecuniary discrepancy between an agreed compensatory damages amount and A's anticipated loss, for the parties are allowed a 'generous margin'.[17] Thus, Anglo-Australian private law provides the parties with a margin of appreciation. Chapter 7 examined three leading theories of the rule. Each failed to explain why such a margin of appreciation exists; or why it is permissible.

I Explanation

Can the margin of appreciation be explained rationally? To address this issue, three steps are necessary.

---

[15] Ernest J Weinrib, *Corrective Justice* (OUP 2012) 94.

[16] John Gardner, 'What is Tort Law For? Part 1: the Place of Corrective Justice' (2011) 30 Law & Phil 1, 39.

[17] *Murray v Leisureplay plc* [2005] EWCA Civ 963, [2005] IRLR 946 [43]; *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [75(iv)].

To start, agreed damages clauses typically deal with low-probability, temporally-distant events. In those cases, the assessment of future loss, gain, or the value of the right inevitably must involve conjecture. Accordingly, usually the estimate will be inaccurate. As Lord Scarman said in *Lim Poh Choo v Camden and Islington Area Health Authority*, 'Knowledge of the future being denied to mankind, so much of the award as is to be attributed to future loss … will almost surely be wrong. There is really only one certainty: the future will prove the award to be either too high or too low.'[18] So, the fact a discrepancy is permitted is at base an inescapable consequence of the human condition. Given the lack of a remarkably powerful crystal ball, the future is unknowable. The fact that it is ordinarily difficult to predetermine the breach or its consequences means that, to be workable, any test of validity *must* provide the parties with a margin of appreciation.

Next, the width of the margin of appreciation is linked to the difficulty of the pre-assessment.[19] That is, the more difficult it is to predetermine the nature of the breach or its consequences, the more latitude is granted. This makes sense. One expects a more accurate forecast of the weather this evening, than the weather in one month's time. However, in most commercial cases, predicting future events is *exceedingly* difficult. Normally, the parties include an agreed compensatory damages clause in their contract precisely because A's anticipated loss is incredibly difficult to predict. It follows that the margin rationally must be 'generous'.

Lastly, the fact the test of validity is applied prospectively, and that in most cases the breach or its consequences are very difficult to predict, means that generally only a very wide gulf will suffice for a court to be *confident enough* (on the balance of probabilities) that the agreed damages clause likely will go beyond the next-next best position, and thus function punitively. Otherwise, the events that the parties have contemplated may in fact occur, such that A will indeed suffer the relevant loss or gain, or be deprived of the value of the relevant infringed right. If so, ultimately the clause simply will fulfil a legitimate function: of compensation, disgorgement or substitution respectively. Thus, consider the case where an agreed compensatory damages amount exceeds A's anticipated loss by a significant margin, but not so much as to make the amount *patently* excessive. If so, a judge could not be sufficiently sure that many of commercial life's vicissitudes would not befall A. If these occurred, the clause would only end up placing A into the next best position by compensating A for her incurred loss. It would do nothing more. In other words, where an exercise involves an element of

[18] [1980] AC 174 (HL) 183.

[19] *Philips Hong Kong Ltd v A-G of Hong Kong* (1993) 61 BLR 41 (PC) 59.

difficult, prospective approximation, only a 'manifestly excessive' standard provides a judge with the requisite level of certainty that the clause is invalid.

II Permissibility

Given this explanation of why a margin of appreciation exists, the question arises as to whether it is permissible on a correctivist view.

According to Weinrib, the same relationship of right and duty moves through a sequence of stages. At each stage, to discharge his duty, a defendant must undertake a specific action. That action is determined by what the duty requires *at that stage*, not at a previous or subsequent stage. Accordingly, the necessary action may differ across the various stages, and the defendant cannot necessarily 'satisfy the duty as it existed at one stage by performing the action called for at a subsequent stage'.[20] What counts is only that for which the duty calls at the stage in question.

So, when a judge assesses the duty to pay agreed damages as at the contract date, for what properly does the duty call? It calls for a valid *taxonomical warrant*. The duty to pay must notionally represent a legitimate response that would come to pass if future events unfolded in a probable, or possible but not far-fetched, manner. If so, it can be said that the clause properly instates the next-next best position.

An example should help. In *Bay Bon Investments Pty Ltd v Selvarajah*,[21] a borrower entered into two loan agreements with a lender. In each case, the principal sum was $300,000; the term was one month; and the same party guaranteed the borrower's obligations. Each contract contained an agreed compensatory damages clause. One clause provided that, upon default, the interest rate on the outstanding sum would increase from 60 per cent to 240 per cent per annum. The other clause stated that, upon default, the rate would increase from 100 per cent to 360 per cent per annum. The borrower defaulted. The lender sought to enforce the agreements against the guarantor. The guarantor claimed the clauses were penal.

White J held that a market for 'loans of last resort' exists that might justify an agreed compensatory damages clause applying a 'very high' default interest rate.[22] The judge then dealt with three separate scenarios.

---

[20] Ernest J Weinrib, *Corrective Justice* (OUP 2012) 93.

[21] [2008] NSWSC 1251.

[22] *Bay Bon Investments Pty Ltd v Selvarajah* [2008] NSWSC 1251 [54].

First, his Honour accepted that a market existed in which the lender likely would have been able to re-deploy its money at very high interest rates, of 60 per cent to 100 per cent per annum. Thus, an agreed compensatory damages amount calculated by applying a default interest rate in this range would have reflected the maximum anticipated loss that the lender would have been *likely* to suffer. Assessed at the contract date, the judge could not know whether matters would unfold as the parties had then anticipated. However, the degree of probability would be high enough to leave the judge confident that the amount may well turn out, in the end, to be compensatory. It follows that a valid taxonomic warrant would exist: compensation. Consequently, the clause would be valid.

Second, White J held that it was possible that a market of 'even 120 per cent per annum' may have existed.[23] So, an agreed compensatory damages amount calculated by applying this default interest rate would have reflected a prospective loss that the lender may have *possibly* suffered, if *some but not all events went against it* in the course of its business. Although the judge could not know how things would ultimately turn out, he could not be sure enough that in the end the amount would not be compensatory. As such, a valid taxonomic warrant would exist. Accordingly, such a clause would be upheld.

Finally, the judge held that he could not accept that a market existed in which the lender could lend at interest rates of 240 per cent or 360 per cent per annum. Put differently, the instant agreed damages amounts, calculated based on these default interest rates, reflected the maximum anticipated loss that the lender *possibly* would suffer if *all events went against it*, based upon far-fetched factual assumptions. As regards each clause, a court could be confident enough that, on the balance of probabilities, events would not unfold in this way; the lender would not suffer this degree of loss; and part of the agreed amount of 'compensation' would be unnecessary. Consequently, the clauses could not be said solely to represent a compensatory response. In White J's words, 'there comes a point … where a rate is so high' that it operates 'as a punishment for default rather than as compensation to the lender for being kept out of its money'.[24] It follows that a valid *and invalid* taxonomic warrant existed. The clauses, therefore, were penal.

In short, for a duty to pay agreed damages to be valid, a court must be able to say that, assessed prospectively and in light of what may well happen, the agreed damages amount represents a legitimate response. In other words, that amount must have a valid taxonomical

---

[23] *Bay Bon Investments Pty Ltd v Selvarajah* [2008] NSWSC 1251 [55].

[24] *Bay Bon Investments Pty Ltd v Selvarajah* [2008] NSWSC 1251 [54].

warrant. If so, then assessed as at the contract date, the duty will satisfy the demands of corrective justice.

In conclusion, a classificatory, corrective justice theory scores extremely well on the fit criterion for assessing a theory of the Penalty Rule.

2 Transparency

A theory is transparent if it conforms to the courts' general reasoning. Chapter 7 showed that the substantive unconscionability theory struggles to pass this criterion, and the efficiency theory flatly fails it.

Compare a classificatory, corrective justice theory. This theory matches the reasons that judges give for their decisions, and the nature of judicial reasoning itself. A correctivist rationale, then, scores very highly on the transparency criterion. Two examples must suffice.

For one, the preventing punishment principle accords closely with what the courts *say* they are doing. As Chapter 1 noted, a theory is strongly transparent if the explanation that a theory gives tightly matches the judges' reasoning in cases in the particular private law area being examined. Recall the judicial statements outlined in Chapter 8.[25] These illustrated that English and Australian judges strongly support the propositions that penalty clauses are punitive in nature, and that Anglo-Australian private law does not enforce penalty clauses because it prevents punishment. In particular, in English law, judges have adopted this view for over a century. It follows that a theory of the rule that explains it in terms of prohibiting punishment is strongly transparent.

For another, the corrective justice theory adheres strictly with what an Anglo-Australian judge *understands* herself to be doing. So, corrective justice theory is cashed out in terms of rights, obligations and remedies. As the theory is rights-based, it is one that is 'internal' to the private law. That is, in rationalising the Penalty Rule, corrective justice theory uses a conceptual apparatus, language, and reasoning that is broadly the same or similar as that which legal actors use; and that is of the same kind as that which courts routinely adopt.[26] Although the correctivist explanation of the rule is more abstract than the judicial one, the former clearly leads to the latter by logical steps. To adopt Posner's words, the theory plainly

[25] Text to nn 65-72 in ch 8.

[26] Stephen A Smith, *Contract Theory* (OUP 2004) 149.

is one 'drawn "from" law' and is 'articulated in lawyers' jargon'.[27] In brief, the corrective justice rationale respects the private law's self-understanding.

Accordingly, on a taxonomic, correctivist view, the Penalty Rule is precisely what the judges purport it to be. Consequently, the propounded theory is the most transparent on offer.

3 Coherence

According to interpretive scholarship, two types of 'coherence' exist. Under the *weak* notion, a theory satisfies coherence to the extent that it shows the private law, or parts of it, to avoid self-contradiction and to be internally consistent. Pursuant to the *strong* notion, a theory is coherent if it shows that the private law, or parts of it, can be understood as a unified system, perhaps under a single, central theme. Two of the prominent theories of the Penalty Rule examined in Chapter 7 did not perform well on this criterion. On the one hand, the substantive unconscionability rationale failed to render contract law either weakly, or strongly, coherent. On the other hand, the efficiency theory was unable to render contract law, or the private law, even weakly coherent.

By contrast, a classificatory, corrective justice theory of the rule scores extremely highly on the coherence criterion, as it satisfies *both* notions of the concept.

A Weak coherence

A classificatory, correctivist theory satisfies the weak notion of coherence, for it reveals the private law to be internally consistent.

To explain, Heydon J has argued extra-judicially that judge-made laws should be 'substantially consistent' and 'harmonious, without anomaly or internal conflict. They should hang together in a connected fashion.'[28] If the principle that properly underpins the Penalty Rule is one against punishment, then the rules of Anglo-Australian private law are congruent and make sense when taken together. Two examples come to mind.

First, if one accepts that the Penalty Rule implements a preventing punishment principle, then the rule is in perfect harmony with the rule against punitive damages for breach

---

[27] Richard A Posner, 'Legal Reasoning From the Top Down and From the Bottom Up' (1992) 59 U Chi L Rev 433, 433.

[28] JD Heydon, 'Limits to the Powers of Ultimate Appellate Courts' (2006) 122 LQR 399, 414. See also *Dietrich v R* (1992) 177 CLR 292 (HCA) 320.

of contract. Rowan proves the point thus. The writer argues that if English contract law attempted to introduce punitive damages for breach of contract, it would be irreconcilable with its stance on penalty clauses. As penalty clauses and punitive damages share a similar theoretical motivation, Rowan argues that 'a rule that the former are invalid but that the latter are not would create inconsistency'.[29] In other words, if Anglo-Australian contract law permitted one remedy but not the other, it would produce inconsistency and thus render the law not even weakly coherent. This is because incoherence is an 'inevitable consequence of … introducing punitive objectives into contract law's framework of corrective justice'.[30]

Second, in *Harris v Digital Pulse Pty Ltd*,[31] the New South Wales Court of Appeal held that punitive damages are unavailable for a breach of fiduciary duty. A main reason for so holding was an argument regarding 'incompatibility'. Spigelman CJ first noted that issues of incoherence had 'emerged as a concern of contemporary jurisprudence'; and then held that one reason for rejecting the power to award punitive damages for breach of fiduciary duty was that 'a punitive monetary award is incompatible with the law relating to penalties …'.[32] Equally, Heydon JA held that 'The proposition that [punitive] damages are not available in equity is supported by … the equitable rules relating to penalties and forfeitures …'.[33]

According to Smith, if a theory shows that the private law contains inconsistent elements, then it makes the law less intelligible.[34] The reverse must also be true: a theory that renders the private law wholly or mostly consistent is more intelligible. This is indeed the case with a classificatory theory.

B Strong coherence

The corrective justice rationale satisfies the strong idea of coherence, for it is able to assist in unifying the private law, or parts of it, under a single theme. The theory is able to produce greater unity in the private law by bringing the Penalty Rule under a high-level norm that attempts to justify the entire private law, or at least large segments of it.

---

[29] Solene Rowan, 'Reflections on the Introduction of Punitive Damages for Breach of Contract' (2010) 30 OJLS 495, 509.

[30] Ernest J Weinrib, *Corrective Justice* (OUP 2012) 182-83.

[31] [2003] NSWCA 10, (2003) 56 NSWLR 298.

[32] *Harris v Digital Pulse Pty Ltd* [2003] NSWCA 10, (2003) 56 NSWLR 298 [57], [61].

[33] *Harris v Digital Pulse Pty Ltd* [2003] NSWCA 10, (2003) 56 NSWLR 298 [337].

[34] Stephen A Smith, *Contract Theory* (OUP 2004) 149.

For many correctivists, all or most of the private law is based on corrective justice.[35] On this view, although each private law rule might be justified at an intermediate level in terms of a different principle, the more abstract corrective justice norm motivates the whole or most of private law. Seen in this light, the 'private law' category displays great conceptual unity. The category is tightly unified under a central concern for attaining justice in interactions between individuals. It neatly stands in contrast to the 'public law' category, which is concerned with achieving justice in interactions between an individual and the community as a collective.

This thesis does not advance the view that all of Anglo-Australian private law is based on, and justified by, corrective justice. But nor does it undermine that view. A lesser argument suffices to make the point advanced. Corrective justice provides the conventional explanation for several categories in the private law, including the law of wrongs in the tortious and contractual contexts, and the law of unjust enrichment. For instance, according to Gardner, in these cases 'correction is called for and the norm that regulates the correction is a norm of corrective justice'.[36] Accordingly, it can be said that, on a traditional view, a *possibility* exists that corrective justice properly explains all of the private law.

Although strong coherence is not mandatory, it is desirable; and hence judges and theorists should continue to strive for it.[37] If the possibility exists that corrective justice justifies all of the private law, then the corrective justice theory of the Penalty Rule displays the *potential* to move the private law closer towards that possibility being realised, and consequently towards adopting a more unified position. Given this, the instant rationale contributes towards the private law, or a part of it, potentially being understood as a unified system. It thus satisfies the strong notion of coherence. Accordingly, the corrective justice theory scores very well when evaluated against a coherence criterion.

The general point is this. A classificatory, correctivist view of the Penalty Rule is best because it renders the private law internally consistent, and strongly coherent in the required sense.

4 Practice

---

[35] cf Ernest J Weinrib, *The Idea of Private Law* (Harvard UP 1995); Ernest J Weinrib, *Corrective Justice* (OUP 2012) (corrective justice underpins the whole of the private law) and Allan Beever, *Rediscovering the Law of Negligence* (Hart Publishing 2007) 71 ('almost all' of the private law is based on corrective justice).

[36] John Gardner, 'What is Tort Law For? Part 1: the Place of Corrective Justice' (2011) 30 Law & Phil 1, 22 (footnote omitted).

[37] Allan Beever, *Rediscovering the Law of Negligence* (Hart Publishing 2007) 24.

Chapter 7 evaluated the substantive unconscionability and efficiency theories of the rule. It showed that both theories raise practical concerns. Further, the preventing 'overcompensation' theory also generates a practical issue. So, if a private lawyer adheres to this traditional rationale, she must use the term 'overcompensation' when explaining the rule to her commercial clients. However, Chapter 7 noted that, when used in this context, the meaning of the term 'overcompensation' does not accord with its ordinary meaning.[38] It follows that the lawyer would not be using the term in the same way that her client would use it. This may not *hinder* understanding, but it certainly will not *aid* understanding.

Although the prominent theories of the rule raise practical problems, this is not the case with the classificatory, corrective justice theory. In fact, it goes the other way.

In common usage 'penalty' means 'A punishment imposed for breaking a law, rule or contract'.[39] Equally, Anglo-Australian judges equate the term 'penalty' with the term 'punishment'.[40] The two notions are intertwined. Accordingly, if private lawyers embrace the preventing punishment principle and the corrective justice norm that it enunciates, it will create an important practical benefit: it will allow commercial clients easily to discern the rule's essence. This will aid commercial understanding. For instance, assume that a lawyer is advising A in the paradigm scenario.[41] If so, she will explain that an agreed damages clause is enforceable if it does not impose a penalty—or a punishment—on B, meaning the agreed damages payment must not go beyond the sum needed to compensate for loss, return or strip gains, or satisfy the value of an infringed right. As the lawyer's use the term 'penalty' marries up with her client's use of the term, this explanation is simple, comprehensible, and practically beneficial.

In short, on the corrective justice view, commercial parties will (rightly) take the word 'penalty clause' to mean precisely what it says.

5 Conclusion

This chapter demonstrated that on each of the relevant criteria for assessing private law theories, the classificatory, corrective justice theory performs better—and in some cases

---

[38] Text accompanying and following n 151 in ch 7.

[39] Oxford Dictionaries <www.oxforddictionaries.com/definition/english/penalty?q=penalty> accessed 12 January 2015.

[40] eg *Robertson v Alexander (Driver's Trustees)* 1881 SC 555 (IH) 562; *Alder v Moore* [1961] 2 QB 57 (CA) 66; *Robophone Facilities Ltd v Blank* [1966] 1 WLR 1428 (CA) 1446.

[41] Text following n 20 in ch 1.

significantly better—than any of the leading theories of the Penalty Rule. When this rights-based theory is evaluated using interpretive criteria, it displays a tight fit, is highly transparent and is powerfully coherent. The evaluative process also revealed that this taxonomical, correctivist rationalisation is as satisfying in practice as it is in theory.

Chapter 7 established that the leading theories of the rule are unsatisfactory. Chapters 8 and 9 have established that a classificatory, corrective justice theory explains the rule rationally. Accordingly, Anglo-Australian law is faced with a stark choice: the Penalty Rule is either based on corrective justice, or it has no justification whatsoever. The former view is more attractive. It is more attractive simply because it is true. Accordingly, the third aim of this thesis, set out in Chapter 1, has been achieved.[42]

A taxonomical analysis has revealed the rule's true basis. That basis has been justified. Now, one must refine Anglo-Australian law to ensure that the rule aligns with its proper theoretical explanation. The next part undertakes that task.

[42] Text following n 3 in ch 1.

Part Three: Refinement

10 RECONCILIATION

Chapter 1 introduced the argumentation strategy of a case law-focused reflective equilibrium. This approach proceeds in three phases. First, it is necessary to examine the relevant case law and derive a principled understanding of the law from it. Accordingly, Chapters 3 to 6 considered the three stages of analytical inquiry under the Penalty Rule. Those chapters outlined the interpretation that casts the rule in its most coherent and appealing light, *based on the present authorities*. Second, one must produce a theory that best explains the law as restated. Consequently, Chapters 8 and 9 revealed the rule's proper theoretical explanation. It illustrated that a classificatory, correctivist theory justifies the law. Finally, it is essential to re-examine the law with that understanding and theory in mind with a view to refining the law if necessary, such as to produce a greater equilibrium between the law and the theory. That is this chapter's mandate. It seeks to reconcile the rule as restated in Chapters 3 to 6 with the theoretical justification elucidated in Chapters 8 and 9, thus casting the rule in the best possible light, *all things considered.*

This chapter is separated into four sections. The first considers a pressing methodological issue. The next three sections correspond to each of the three stages of inquiry. Accordingly, the second section presents a refined jurisdictional test; and the third elucidates a refined test of validity. The final section considers the consequences of a court validating, and invalidating, an agreed damages clause pursuant to the realigned tests.

The chapter concludes that, viewed in the best possible light, the Penalty Rule invalidates penalty clauses. A penalty clause should be interpreted as a provision that, upon a party's default, imposes a detriment that is manifestly excessive in light of the provision's legitimate function. The rule invalidates a penalty clause by rendering it partially unenforceable. The other party, however, may partially enforce the provision to the extent of its validity. Further, the chapter illustrates that the refined rule produces results consistent with the decisions made in most of the cases, but arrives at those results through more perspicuous and forceful reasoning.

This chapter espouses a refined rule that is lucid, systematic and coherent. Most importantly, it adjusts our understanding of what the rule *actually* requires so as better to serve its true justification. It thus achieves equilibrium between the penalty doctrine as we find it, and the best explanation of that doctrine.

1 Methodology

This chapter strives to refine the Penalty Rule in the light of its true theoretical basis. Like the restatement process, the refinement process requires an interpretive approach. However, in this context, the emphasis differs. As Chapter 1 noted, the interpretive framework contains historical, descriptive and normative elements. Although when refining a rule one is still seeking how best to understand the law, in undertaking that process the normative element of an interpretive approach is more pronounced. Put differently, refinement is more about what the law ought to be rather than what it is; yet what it ought to be is heavily constrained by the history and current shape of the case law.

Accordingly, legislation aside, only an ultimate appellate court could undertake a refinement process. That said, an ultimate appellate court judge could easily enunciate the refined test that this chapter proposes. As illustrated below, the improved test is one that she might fashion by applying orthodox judicial methods, and taking only small interpretive steps.

2 Jurisdiction

This section considers the jurisdictional test. It begins by presenting a refined jurisdictional test. It then investigates the most problematic issue arising from this test: namely, how to identify a substantive breach of contract. Lastly, this section applies the realigned jurisdictional test to difficult fact patterns arising from the cases and the literature, illustrating how the test will operate in commercial practice.

A The test

In order to arrive at a refined jurisdictional test, three steps are required.

First, one must outline the jurisdictional test as restated. Thus, Chapter 4 concluded that, properly understood, the jurisdictional test contains two requirements: B must suffer a *detriment*, and B's *breach* of contract must trigger that detriment. As to the former requirement, a relevant detriment will exist if B comes under any of the following duties: to pay money to A; to transfer a personal or proprietary right to A; or to extinguish a right held against A that obligates A to transfer money or property to B. As to the latter requirement, B's breach of contract provides the 'core case' in which B may invoke the court's review power. However, the review jurisdiction includes cases where B's detriment occurs due to an event constituting a causal or substantive breach. Accordingly, the jurisdiction properly extends to a default, or a causal breach.

Second, we must consider again the rule's theoretical explanation. So, Chapters 8 and 9 revealed that a classificatory, correctivist theory justifies the Penalty Rule. The two-level theory provided two relevantly important insights. The classificatory theory revealed that the breaches in play are consent-based, not wrongs-based. Also, the corrective justice theory explained why the actual breach requirement exists, and rationalised why the rule's scope properly extends to substantive breaches. In short, a consent-based actual wrong violates the mandatory norms of corrective justice, and a consent-based substantive wrong attempts to stultify those norms.

Finally, one must refine the scope of the restated jurisdictional test, in light of the law's correct theoretical justification. Accordingly, cast in the best possible light, the review jurisdiction only arises where B incurs a *detriment* as a result of B's *default*.

As to the first limb, for Anglo-Australian courts to take jurisdiction, the clause must cause B to suffer a relevant detriment. Three types of obligations suffice to satisfy this requirement. To start, the clause obliges B to pay a sum of money to A. Next, the provision imposes upon B a duty to transfer a personal or proprietary right to A. Lastly, the clause obliges B to extinguish a personal right B holds against A.[1]

As to the second limb, to invoke the court's review power, B's default must be the trigger for him to incur the relevant detriment. The default may be a consent-based actual breach, or a consent-based substantive breach. Further, the notion of a default includes a 'causal default': that is, a default that is the causally operative trigger for B's detriment.

B Breach

The refined jurisdictional test includes within its scope a consent-based *substantive* breach of contract. This concept is linked to the notions of a 'disguised' agreed damages clause and a 'disguised' penalty clause. As Chapter 4 discussed, a 'disguised' agreed damages clause refers to a clause drafted in an attempt to evade—and hence stultify—the jurisdictional rule. If such a clause is invalid, it is known as a 'disguised' penalty clause. A substantive breach of contract is the event triggering the relevant detriment under a 'disguised' agreed damages clause.

---

[1] Note that, in rare cases, it is possible that B may be obliged to extinguish a *proprietary* right: see n 25 in ch 4.

In *Euro London Appointments Ltd v Claessens International Ltd*,[2] Chadwick LJ described the nature of a 'disguised' agreed damages clause. Such a clause involves attempting to present the event triggering the relevant detriment as something that truly it is not. Further, his Lordship held that if the 'true nature of the bargain' is such that it involves a 'contractual obligation',[3] then a court will treat the clause on that basis: that is, as involving a substantive obligation. If a party violates a substantive obligation, it will result in a substantive breach.

The phrase 'substantive breach' serves only as an outline. It does not sufficiently distinguish cases where the parties have attempted to circumvent the jurisdiction from cases where the parties have entered into a genuine commercial agreement. This raises the 'central difficulty'[4] for the refined jurisdictional test: how are Anglo-Australian courts practically to identify a substantive breach of contract? We examine that issue now.

I The 'core case'

To develop a test for identifying a substantive breach of contract, it is best to work by analogy from a 'core case'. This technique provides the great engine room for common law development. It is a method adopted by judges and jurists alike.[5]

As Chapter 4 illustrated, a clause activated by an actual breach of contract provides the 'core case' where the court's review jurisdiction arises. An actual breach of contract, in turn, exists where B breaches *an obligation* to perform. Proceeding analogically from this 'core case', the question arises: in this context, what are the distinctive features of an actual obligation? If we identify the relevant indicia of an actual obligation, we can then examine B's relationship to a particular act or event to see whether it bears those indicia. If it does, then B's relationship to that act or event is sufficiently analogous to an actual obligation to label accurately B's relationship a 'substantive obligation', capable of leading to a substantive breach.

In this context, there are three defining or characteristic features of an actual obligation.

---

[2] [2006] EWCA Civ 385, [2006] 2 Lloyd's Rep 436.

[3] *Euro London Appointments Ltd v Claessens International Ltd* [2006] EWCA Civ 385, [2006] 2 Lloyd's Rep 436 [28].

[4] *Andrews v Australian and New Zealand Banking Group Ltd* [2011] FCA 1376, (2011) 211 FCR 53 [74], revd [2012] HCA 30, (2012) 247 CLR 205.

[5] eg Peter Birks, *Unjust Enrichment* (2nd edn, OUP 2005) ch 1; HLA Hart, *The Concept of Law* (3rd edn, OUP 2012) 81.

To start, if a party is obligated to do an act, necessarily she will be responsible for that act. The party is responsible both in the sense that she is bound to ensure that the relevant event occurs, and is accountable for the consequences of it failing to occur.

Next, if a party undertakes a duty to perform, inevitably she will be able to exercise control over whether she performs or not. Conversely, if a party cannot control whether performance occurs (such as whether an event takes place), it does not make sense to say that she is 'obligated' to perform. For instance, one cannot be obligated to ensure that it does not rain in London tomorrow.

Lastly, an obligated party has no real choice as to whether to undertake the relevant act or not. In other words, she is compelled to act. If a party can freely choose to act otherwise, then the situation lacks the necessary degree of compulsion to elevate the required act to the status of a 'duty'.

Accordingly, three pertinent indicia of an actual obligation exist: it places the obliged party under a *responsibility* to ensure an event occurs; the party can *control* whether that event in fact occurs; and the party is *compelled* to ensure that the event occurs. The second and third indicia are best viewed as subsets of the first: a party only has a responsibility to ensure that the relevant event occurs if she can exercise control, and is compelled to act.

Thus, working by analogy, a substantive obligation will also display these features. However, the nature of the 'compulsion' will change given the new context. In the actual obligation context, the compulsion is *legal*. Conversely, in the substantive obligation scenario, given that per hypothesis a legal obligation does not exist, the compulsion can be *practical* only. Consequently, a substantive obligation can only exist where a party is placed under a degree of practical compulsion to perform that is the same as, or similar to, the degree of legal compulsion to perform existing in the case of an actual duty to perform.

To identify a substantive breach, then, Anglo-Australian courts must apply the following test. First, one must ask: is B substantively *responsible* to ensure that an event occurs, or does not occur? Crucially, B is substantively responsible for an event if B can *control* whether the event occurs, or does not occur; and the discrepancy in consequence between the event occurring, and it not occurring, is sufficient *practically to compel* B to ensure that the event occurs, or does not occur. If the answer to the question put is 'yes', then a substantive obligation exists. If so, one must then ask: has B incurred the relevant detriment as a result of the occurrence or non-occurrence of the event for which B was substantively responsible? If the answer to this question is 'yes', then a substantive breach exists. Therefore, the clause is reviewable.

II Responsibility

The proposed test contains an overarching concept of 'responsibility'. Does the Anglo-Australian case law support the use of such a concept?

As Chapter 4 noted, two Australian judges have held that the jurisdiction properly extends to a substantive breach, and that the appropriate test to identify such a breach centres on the responsibility notion—that is, Deane J in *AMEV-UDC Finance Ltd v Austin*;[6] and Brereton J in in *Integral Home Loans Pty Ltd v Interstar Wholesale Finance Pty Ltd*,[7] in which his Honour espoused the Responsibility Test.[8]

At first blush, the Australian High Court's decision in *Andrews v Australia and New Zealand Banking Group Ltd*[9] appears to reject any responsibility-based test. The High Court held that a clause may engage the Penalty Rule jurisdiction even if B has 'no responsibility or obligation' to avoid the occurrence of the event triggering the detrimental obligation.[10] One might read the Court as saying that a clause is reviewable even if the detriment-triggering event is *not* one for which B was responsible. If that reading is right, then the proposed test is interpretively problematic.[11] However, another reading is available. The judgment may be interpreted as one where the Court used 'responsibility' as a synonym for 'obligation', attempting merely to emphasise that a breach of a contractual obligation did not delimit the court's review jurisdiction. If that reading is correct, then the proposed approach remains open.

For the following three reasons, the latter interpretation is the better one.

First, the judgment must be read in light of the arguments put to the Court. The applicants' Counsel had submitted that a 'penalty' was properly defined as 'an obligation to pay money or transfer property which is conditioned on the occurrence or non-occurrence of an

---

6 (1986) 162 CLR 170 (HCA) 199.

7 [2007] NSWSC 406, (2007) 2 BFRA 23 [71], [74], revd [2008] NSWCA 310, (2008) 257 ALR 292.

8 Text to n 47 in ch 4.

9 [2012] HCA 30, (2012) 247 CLR 205.

10 *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [84]. See also [78].

11 In particular, the *stare decisis* doctrine may present an obstacle: see JD Heydon, 'Limits to the Powers of Ultimate Appellate Courts' (2006) 122 LQR 399, 403.

event placed within the … responsibility of [B]'.[12] The applicants were successful in the appeal. One must infer that the High Court decided the case on the submissions as put, and accepted this contention.

Second, as Chapter 4 noted, the High Court lauded Brereton J's judgment in *Interstar*. In particular, the Court referred to and cited with approval the passage of the *Interstar* judgment containing the Responsibility Test.[13]

Finally, consider the first available interpretation. It suggests that a clause may engage the rule even though B is not responsible or obliged to ensure that the relevant event occurs, or does not occur, such that B avoids incurring a detriment. If that reading is correct, many clauses bearing no resemblance whatsoever to the 'core case' will fall within the review jurisdiction. Per hypothesis, a breach of contract will no longer form the 'core case'. Accordingly, two interpretations of the judgment are open: one will result in an incremental, analogical extension of the settled 'core case'; and the other will explode it. Rationality, legal certainty and orthodox judicial methods favour the former.

III Control

Under the proposed test, B is substantively responsible for an event only if B can control whether the event occurs, or does not occur. Thus, B will commit a substantive breach only if he has the power *significantly to influence* whether the circumstance triggering the detrimental obligation does, or does not, come to pass.

Control provides a principled basis for determining whether a clause falls within the review jurisdiction, or not. As best understood, several cases support this proposition. Consider the following two situations.

On the one hand, if B *cannot* influence whether the relevant detriment-triggering event occurs, or fails to occur, then the clause is unreviewable. A case best explained on this basis is *Philip Bernstein (Successors) Ltd v Lydiate Textiles Ltd*.[14] A finance company and a dealer entered into a 'recourse' agreement. The dealer agreed to procure offers from hirers to enter into hire-purchase agreements with the finance company on a commission basis. An agreed damages clause stated that if a hirer defaulted in paying any instalments due under a hire-purchase agreement, the dealer would pay to the finance company the balance of the instalments that

[12] *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205, 207.

[13] *Andrews v Australia and New Zealand Banking Group Ltd* [2012] HCA 30, (2012) 247 CLR 205 [49].

[14] (CA, 26 June 1962), sub nom *Sterling Industrial Facilities v Lydiate Textiles Ltd* (1962) 106 SJ 669.

would have been payable if the agreement had run its course. A hirer defaulted. The finance company sought to enforce the clause. The dealer argued the clause was penal. The English Court of Appeal held the clause was unreviewable. The detriment-triggering event was the hirer's default. However, the dealer and the hirer were at arm's length. Accordingly, the dealer could not influence the hirer's behaviour. In other words, the dealer could not control whether the hirer defaulted.

On the other hand, if B *can* significantly influence whether the relevant detriment-triggering event comes to pass, or not, then the clause is potentially reviewable. A case in point is *Interfoto Picture Library Ltd v Stiletto Visual Programmes Ltd*.[15] So, Interfoto supplied photographic transparencies to Stiletto. It did so on conditions requiring Stiletto to return the transparencies within 14 days from the delivery date, otherwise a 'holding fee' of £5 per day would be charged for each transparency retained beyond that period. Stiletto returned the transparencies well after the deadline. Interfoto sought to recover the holding charge of approximately £3,780. In obiter, Bingham LJ intimated that the clause was potentially 'challengeable as a disguised penalty clause'.[16] Obviously, Stiletto completely controlled whether it returned the transparencies late, or not.

IV Practical compulsion

Pursuant to the refined approach, B is substantively responsible for an event only if the discrepancy in consequence between the event occurring, and not occurring, is sufficient practically to compel B to ensure that the event occurs, or does not occur. Hence, this element of proposed test turns on the fundamental ideas of 'discrepancy' and 'compulsion'.

As to the former, two comments are necessary. For one, it is important to note the relevant 'discrepancy' in play. The question is whether a relevant discrepancy exists between the consequences for B of event occurring and it not occurring, or between B's putative options. For another, the discrepancy between B 'options' must be such that any apparent choice is, in a practical sense, empty. An example: a clause provides B with the 'option' of leasing a chattel for a period up to 1 January for £5,000; or for a period up to 2 January for £20,000. On its face, the difference between the alternatives is so great that B has no real

---

[15] [1989] QB 433 (CA).

[16] *Interfoto Picture Library Ltd v Stiletto Visual Programmes Ltd* [1989] QB 433 (CA) 445-46.

choice. Such a clause would constitute 'a colourable attempt to levy a penalty'.[17] Thus, a court will be prepared to recharacterise it accordingly.

As to the latter, a clause will only satisfy this element of the refined test if it creates the requisite degree of practical compulsion. It must be a degree that is 'obligation-like', in that the clause must create a level of practical compulsion that is the same as, or similar to, the level of legal compulsion that a clause imposing an actual obligation would generate. In other words, B must be practically forced to act.

This strategy is used elsewhere in the private law. So, where a defendant breaches a contract, a claimant may apply for a prohibitory injunction. This injunction acts negatively, by ordering the defendant *not* to do something. It may be contrasted with a specific performance decree, which acts positively, by requiring the defendant *to do* something. It has been held that if a prohibitory injunction restraining the breach of a negative duty practically will compel the defendant to perform a positive duty with the claimant, then an Anglo-Australian court will apply specific performance principles to decide whether to grant or refuse the injunction.[18] In other words, where an application for a prohibitory injunction substantively constitutes a claim for specific performance, a court will treat the action as taking the latter form.

This approach is analogous to the approach proposed here. To illustrate: assume that B has three ways in which to act, being X, Y and Z. In the injunction scenario, if a court legally restrains B from doing X and Y, it will practically force B to do Z. In substance, then, B will be obligated to do Z. In the Penalty Rule jurisdiction scenario, if the discrepancy in result between the 'options' makes doing X and Y commercially unrealistic or financially prohibitive, it practically forces B to do Z. In substance, B is obliged to do Z. Thus, the degree of practical compulsion to undertake a particular act provides a guide to determining whether a substantive obligation will or does exist respectively.

The use of a case example may help. As Chapter 4 discussed, one common avoidance technique used to create a 'disguised' agreed damages clause is the drafting of a clause that prima facie appears to be option clause. These putative option clauses give B no real, practical choice as to whether to perform, given the consequences attached to B's 'choosing' one 'option' rather than another. Accordingly, these 'option' clauses display the requite degree of practical compulsion.

---

[17] Michael Bridge, *The Sale of Goods* (3rd edn, OUP 2014) [12.13] (footnote omitted).

[18] *Warren v Mendy* [1989] 1 WLR 853 (CA) 865-66, 867; *Atlas Steels (Australia) Pty Ltd v Atlas Steels Ltd* (1948) 49 SR (NSW) 157 (NSWSC) 163, 166 cf *LauritzenCool AB v Lady Navigation Inc* [2005] EWCA Civ 579, [2005] 1 WLR 3686 [12], [19]-[22], [33].

Consider now *The General Trading Co (Holdings) Ltd v Richmond Corp Ltd*.[19] The General Trading Company agreed to purchase shares from Richmond. The General Trading Company argued that the sale agreement should be construed as a contract to purchase shares at two alternative prices: namely, £600,000 where Richmond procured a guarantee; and £60,000 where Richmond did not so procure. The disproportionality in result between the two 'options' was likely to force Richmond to procure the guarantee. Put another way, the offer of a 'choice' was, in a practical sense, empty. Richmond was thus effectively compelled to act. Given Richmond exercised control over whether the guarantee was procured, it was substantively responsible to ensure that the guarantee was obtained. Therefore, on this interpretation, the clause constituted a reviewable 'disguised' agreed damages clause. Notably, this outcome accords with the Court's decision that the relevant clause engaged the review jurisdiction.[20]

C Applied

The true scope and proper application of the refined jurisdictional rule is best illustrated through several examples. These examples show how the equilibrial rule would generate the same outcome as that reached in the leading cases, but would attain that result through more elegant and unified reasoning. They also demonstrate that the proposed test is able to operate effectively in 'hard' cases.

The refined jurisdictional approach provides the most elegant and appropriate method of reaching the court's preferred conclusion in many cases. Take those cases involving a putative condition precedent clause, for instance. A condition precedent clause provides that a party's right will not accrue until a certain event occurs, or fails to occur. For example, a clause may state that X will not obtain a right to be paid until she has completed her work for Y. English and Australian appellate courts have held that some types of condition precedent clauses are unreviewable. Properly understood, this is because the relevant clauses do not produce the requisite degree of practical compulsion. We consider each jurisdiction in turn.[21]

---

[19] [2008] EWHC 1479 (Com Ct), [2008] 2 Lloyd's Rep 475.

[20] Beatson J held that the agreement properly could not be construed as an option contract. Rather, the agreement contained an agreed damages clause activated by a breach of obligation, which was clearly reviewable: *The General Trading Co (Holdings) Ltd v Richmond Corp Ltd* [2008] EWHC 1479 (Com Ct), [2008] 2 Lloyd's Rep 475 [114]-[118].

[21] Not all clauses containing a condition precedent will fall outside the refined review power. Thus, it is likely that a clause imposing an entire obligation as to quality will satisfy the refined test, and will engage the review jurisdiction. For example, see the facts of *Eshelby v Federated European Bank Ltd* [1932] 1 KB 423 (CA). In that

We begin with the 'central' English case of *Euro London*. An employment agency entered into a contract with a client. The agency agreed to find suitable applicants for the client to employ, and to arrange introductions. If the client engaged the applicant, it was to pay a fee. A clause stipulated that the client was entitled to a refund of a proportion of the fee if an applicant's employment was terminated within 12 weeks. A further provision stated that, in order to qualify for a refund, the client had to pay the agency's fee within seven days of the invoice date. The agency introduced an applicant to the client. The client employed the applicant, but failed to comply with the qualifying provision. The applicant left in less than 12 weeks. The client sought a partial refund. The agency refused. The client argued that the qualifying provision was penal. The English Court of Appeal held that the provision was simply a condition precedent to any right to a refund, and did not fall within the review power.[22]

As best interpreted, the provision contained a non-promissory condition precedent in the nature of a qualified privilege: that is, a special benefit granted 'over and above' a person's strict right. These clauses do not practically compel performance in the same way as a clause that imposes an actual obligation. A party will have a complete discretion regarding whether to fulfil the relevant qualifications to obtain the benefit, or not. Accordingly, it cannot meaningfully be said that the party is 'responsible' to ensure that the stated conditions are met, and the benefit obtained. On the facts of *Euro London*, then, no substantive obligation existed, such that no substantive breach could have occurred. Thus, the clause was properly unreviewable.

Now consider the important 'standard' Australian case of *Helicopters Pty Ltd v Bankstown Airport Ltd*.[23] Bankstown subleased part of an airport to Helicopters. The sublease obligated Helicopters to pay Bankstown the outgoings. It also entitled Bankstown to review the rent every two years. Pursuant to clause 4.1(e), Helicopters only obtained the right to object to Bankstown's rent review notice if it had paid all moneys then owing under the sublease. Bankstown initiated a rent review; Helicopters objected to its rent review notice. However, at that time, Helicopters owed Bankstown $7,756.48 for outgoings. Helicopters argued that clause 4.1(e) was penal. The New South Wales Court of Appeal rejected that argument. It held that clause 4.1(e) properly involved a condition precedent to a contractual right, and did

---

case, however, the penalty point was not pleaded. It follows that whether such a clause is properly reviewable remains undecided cf Edwin Peel, *The Law of Contract* (13th edn, Sweet and Maxwell 2011) [20-146].

22 *Euro London Appointments Ltd v Claessens International Ltd* [2006] EWCA Civ 385, [2006] 2 Lloyd's Rep 436 [19]-[23].

23 [2010] NSWCA 178, (2010) 15 BPR 28,593.

not require payment of an additional amount upon breach. Accordingly, the clause was not within the rule's jurisdictional scope.[24]

Did the discrepancy between the two outcomes practically compel Helicopters to pay? The key to answering this question lies in noting that, like the test of validity, the jurisdictional test is applied *at the contract date*. Accordingly, the two relevant consequences to compare are as follows. On the one hand, if Helicopters paid its total outstanding debts (which might include outstanding rent and outgoings), it would be able to object to Bankstown's review notice. On the other hand, if Helicopters did not satisfy those particular debts, it would not be able to object.

Now, assessed as at the contract date, the difference between paying any outstanding debt and obtaining the right to review, and not paying the outstanding debt and forgoing the review right, may or may not have practically compelled payment. However, if the disputed total debt was more than or equal to the likely potential benefit arising from a review, then little or no practical compulsion would exist. Given that Helicopters was leasing part of an airport—where rent and outgoings were likely to be substantial—this was entirely possible. Accordingly, it followed that the difference between the two consequences would not likely be so great that Helicopters in effect would be driven to pay. Given this, it could not be said that clause 4.1(e) satisfied the practical compulsion element of the refined test.

Further, the proposed jurisdictional test can function successfully in 'hard' cases. That is, the test displays the capacity robustly to differentiate in difficult cases between genuine contract terms and 'disguised' agreed damages clauses. One such case is that of an apparent incentive payment clause providing a bonus for early completion. Recall the following example, given in Chapter 4.[25] Assume that C agrees to construct a building for D. The contract contains a clause setting out the following regime: if C completes the building on or before 1 January, D will pay £1,000,000; if C completes in the period of 2 January to 1 February, D will pay £950,000; if C completes in the period of 2 February to 1 March, D will pay £900,000; and so on. Assuming the requisite detriment exists, does the refined jurisdictional test catch this clause?

Consider first the control requirement. Though more facts would be necessary to determine the matter either way, on the facts as stated seemingly C would have the power significantly to influence whether it would complete construction by any particular date. On

---

[24] *Helicopters Pty Ltd v Bankstown Airport Ltd* [2010] NSWCA 178, (2010) 15 BPR 28,593 [37], [41].

[25] Edwin Peel, 'The Rule Against Penalties' (2013) 129 LQR 152, 154.

its face, then, one cannot say that the detriment is triggered by an apparent non-promissory event outside of C's control.[26]

Whether the practical compulsion requirement is satisfied is difficult. It all depends. On the one hand, it might be argued that the true contract price is £1,000,000. If so, the clause effectively holds that C will forgo the right to be paid five per cent of the price for each month it delays. On the other hand, it might be contended that the true contract price is, for example, £500,000. If so, the clause in reality holds that C will be paid five per cent above the price for each month it completes early, up to a maximum of 10 months.

Ultimately, a court can only decide this matter based on the objective market evidence of contract prices for constructing a building of a similar type.

Thus, if the evidence reveals that construction companies typically charge in the vicinity of £1,000,000 to erect a comparable building, a court likely will prefer the former characterisation. Given this, the discrepancy between performing by 1 January and 1 February, or a later date, appears substantial. In practice, then, the clause probably will force C to deliver up a completed construction by 1 January. Accordingly, C is practically compelled to ensure that it completes by that date.

Conversely, suppose it is proved that organisations usually charge around £500,000 to construct the same type of building. In that case, a judge likely will opt for the latter classification. Given this, the discrepancy between performing by 1 January and, say, 1 February, is not substantial, given that one is only dealing with true *additional* profit. It follows that, for at least the initial 10 months of the project, the clause provides C with a very real choice. Consequently, no practical compulsion exists to ground a substantive obligation.

In conclusion, Harder has lamented the fact that, to date, courts and commentators supporting an 'extended' jurisdiction have failed clearly to define the event that ought to trigger a review under the Penalty Rule.[27] This section has answered the call.

3 Validity

This section examines the test of validity. It first frames a refined test of validity. The section then explains in detail the two central elements of that test: the concepts of 'legitimate

[26] cf Law Commission, *Penalty Clauses and Forfeiture of Monies Paid* (Law Com WP No 61, 1975) [26] ('delay caused by bad weather').

[27] Sirko Harder, 'The Relevance of Breach to the Applicability of the Rule Against Penalties' (2013) 30 JCL 52, 58.

function' and 'manifest excess'. Finally, it systematically applies the improved test of validity to a set of important cases, demonstrating how the test is to function in commercial practice.

A The test

Three stages are required to develop a refined test of validity.

To start, we must reconsider the test of validity as restated. Chapter 5 examined several tests potentially available to an Anglo-Australian court to help it to decide whether a clause is penal. Relevantly, that chapter concluded as follows.

First, the Extravagance Test is the sole test of validity. Anglo-Australian case law has shifted such that as a matter of precedent the Extravagance Test is now *the* test of validity. Further, the test is effective and principled, and exhibits great explanatory power.

Second, the Impossibility Test is dysfunctional if one adopts a traditional, loss-based approach. The test can be saved, however, if one takes a rights-based view. From that perspective, the Impossibility Test merely confirms that the Penalty Rule rightly applies to agreed substitutive damages clauses, and that courts adopt the Extravagance Test to determine the validity of such clauses. The test is best understood on that basis.

Finally, the Indiscriminate Cause Test is avoided or evaded as much as it is applied. This raises a difficult problem regarding how best to interpret the test.

Next, we must reconsider the rule's theoretical justification. Thus, Chapter 8 established that the Penalty Rule implements a classificatory principle of preventing consensual, civil punishment. The preventing punishment principle, in turn, enunciates what corrective justice demands in this context. Relevantly, corrective justice requires this negative principle apply because punishment is an unjustifiable response in the law of consent. Put differently, where the causative event is consent, punishment legitimately cannot be the function of a right that arises in response to it. As it was put in *Andrews*, a penalty is a detrimental obligation that 'has the purpose or character of … punishment. It serves no legitimate function within the parties' relationship …'.[28]

Lastly, one must refine the contours of the restated test of validity, in light of the rule's proper theoretical explanation. It follows that, cast in the best possible light, the test of validity requires an Anglo-Australian court to determine whether the agreed damages clause imposes a detriment that is *manifestly excessive* in light of the clause's *legitimate function*. To identify such a

[28] [2012] HCA 30, (2012) 247 CLR 205, 208 (Counsel's submissions).

clause, the court must apply a comparative test based on the particular legitimate function in play. The test is applied as at the contract date.

B Legitimate function

The refined test of validity requires an Anglo-Australian judge to examine an agreed damages clause and identify the legitimate function that the stipulated sum will perform. Three points arise.

First, presently the most common legitimate function is compensation. However, this thesis has illustrated that other legitimate functions include substitution, restitution and disgorgement. That said, the list of legitimate functions is not closed. The extended Birksian grid set out in Table A[29] demonstrates this, as it contains a header in the 'Responses' rows entitled 'Other functions'.

Second, all agreed damages amounts perform a legitimate function. A liquidated damages sum will perform that function alone. Conversely, a penalty will perform *both* a legitimate function and an illegitimate—punitive—function. This follows from the nature of civil punishment. As Chapter 1 noted, in the private law a punitive payment is one that is parasitic upon, and excessively 'over and above', another payment made which manifests a different response.

An example will clarify the point. Suppose a contract contains a provision merely stating that if B breaches his obligation to do X by Wednesday, B is obliged to pay £500 to A. The first question the court must ask is: how will the agreed damages sum work? Will it—at least partially—compensate A for her loss, substitute for the value of A's infringed right, or strip B of his gain? In some cases, this will be obvious. In others, it may not. Assume that the court holds that the clause is properly an agreed *compensatory* damages clause. Now compare the following two scenarios. On the one hand, the court validates the clause. If so, the liquidated damages sum of £500 wholly performs a legitimate compensatory function. On the other hand, the court invalidates the clause, holding that the penal sum contains an excessive component of £200. Given this, the enforceable element of £300 performs a legitimate compensatory function; the unenforceable part of £200 is punitive, and illegitimate. As such, the clause's essential function is both 'to compensate the non-defaulting party [and] to punish the defaulting party'.[30]

---

[29] Table following n 57 in ch 1.

[30] *CRA Ltd v NZ Goldfields Investments* [1989] VR 873 (VSC) 875.

Finally, an Anglo-Australian court must ascertain the clause's legitimate function *objectively*. As Allsop P stated in *Fermiscan v James*, a court must seek to ascertain the clause's 'true operative character'.[31] Importantly, the issue is whether *the clause* objectively performs a particular function, not whether *one party or both parties* subjectively intended the clause to perform a particular function. Thus, either party's actual intention is largely irrelevant to determining the clause's legitimate operation.

One final issue remains: what is the true status of the Indiscriminate Cause Test? Chapter 5 noted that, in principle, Anglo-Australian appellate courts continue to support the Indiscriminate Cause Test.[32] However, in practice, courts find ways to sidestep the test, applying it selectively. The chapter concluded that Anglo-Australian courts should either jettison the test, or reinterpret it so that its true scope of application is made clear, and then demand that it be respected.

The best way forward is this. As properly comprehended, the Indiscriminate Cause Test requires a *series* of applications of the Extravagance Test. For instance, if an agreed compensatory damages clause applies to 20 breaches, in theory the court must apply the Extravagance Test to each of those breaches to ascertain whether the anticipated loss flowing from any one of them is disproportionate to the agreed damages sum. Accordingly, the Indiscriminate Cause Test is merely a particularised instance of the sole test of validity. It therefore should be faithfully applied. However, it is necessary to clarify the Indiscriminate Cause Test's proper scope of application, and its true effect. Three points follow.

To start, the Indiscriminate Cause Test cannot apply to an agreed substitutive damages clause, for there is only one right to performance. Consequently, courts must only apply it to agreed damages clauses involving anticipated losses and gains (that is, the consequences of a breach), and not clauses involving the anticipated value of the infringed performance right.

Next, as it pertains to losses and gains, this test merely involves applying the Extravagance Test several times. As demonstrated below, rightly understood the test of validity only requires a court to take account of the *likely* consequences, and the *possible but not far-fetched* consequences, of a breach. Applying the Extravagance Test only to breaches with such consequences will narrow the scope of Indiscriminate Cause Test, ensuring that it operates in a commercially acceptable fashion.

---

[31] [2009] NSWCA 355, (2009) 261 ALR 408 [153].

[32] *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [73]-[74]; *Fermiscan v James* [2009] NSWCA 355, (2009) 261 ALR 408 [149]-[152]; *Zachariadis v Allforks Australia Pty Ltd* [2009] VSCA 258, (2009) 26 VR 47 [158].

Lastly, the Indiscriminate Cause Test effectively involves several discrete applications of the Extravagance Test. Given this, if the relevant clause fails *any* application of the latter test, it is incorrect to say that a rebuttable presumption of penalty is raised. In truth, the clause is simply invalid.

C Manifest excess

After the court has ascertained the legitimate function that the agreed damages sum will perform, it can then apply the appropriate comparative test to determine whether the sum contains an excessive component, and hence partially performs an illegitimate punitive function. Given the stated legitimate functions, under the refined comparative test a court must compare the agreed damages sum with the amount of A's anticipated loss or gain, or the anticipated value of A's infringed performance right.[33] For the clause to be valid, the two figures must appropriately correspond. But what is the best way to describe the acceptable level of discrepancy between these two figures?

Presently, Anglo-Australian case law holds that the Extravagance Test is the sole test of validity. Pursuant to that test, an agreed damages sum is in fact penal if it is 'extravagant and unconscionable in amount'[34] compared to A's anticipated loss. Putting to one side the need to expand the responses to which the test applies, classically the terminology courts have used to describe the appropriate degree of discrepancy is that of 'extravagance'.

Chapter 9 established that the rule's theoretical justification requires that the test of validity provide the parties a margin of appreciation. The classificatory, correctivist theory explained that only a standard that encapsulates a wide margin of appreciation will be realistic and workable, and will provide a judge with the requisite level of certainty that an agreed damages clause is actually invalid. That is the best standard that we can adopt, in a world without magic.

What is the best terminology to describe the acceptable degree of discrepancy, whilst encapsulating the wide margin of appreciation that the private law justifiably provides here? It is this. The court must determine whether the agreed damages clause imposes a detriment that

---

[33] This is the correct comparative test where the event that triggers the allegedly penal obligation is an *actual* breach. Conversely, where the triggering event is a *substantive* breach, the court must compare the agreed damages amount with the anticipated loss or gain that will likely result from the event constituting the substantive breach, or the anticipated value of the substantively infringed performance right.

[34] *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd* [1915] AC 79 (HL) 87.

is *manifestly excessive* when compared with the amount of A's anticipated loss or gain, or the anticipated value of A's infringed performance right.

I Case law and principle

Anglo-Australian judges at the highest level have supported the improved 'manifest excess' terminology. For example, in *Photo Production Ltd v Securicor Transport Ltd*, Lord Diplock held that a penalty is a sum that is 'manifestly … in excess of the amount which would fully compensate' A for her anticipated loss.[35]

Further, although the proposed formulation changes the present nomenclature in three respects, each amendment is principled.

First, the recast formulation excises the word 'unconscionable'. Chapter 5 concluded that the expression 'unconscionable' in the Extravagance Test is best interpreted as being a synonym for 'extravagant'. Thus, the term has no independent content. Given this, keeping the expression can serve no useful purpose. In fact, it may confuse and mislead. In particular, if the test of validity retains the term 'unconscionable', it may (wrongly) continue to suggest to some that substantive unconscionability is the rule's proper justification.

Second, the refined terminology jettisons the term 'extravagance' in favour of 'excessive'.[36] The latter term is appropriate in this context because, like the concept of 'extravagance', the very notion of 'excessiveness' moves in one direction only. It thus focuses a court's attention on agreed damages sums that are disproportionately *large*.[37] However, 'excessive' is a preferable term to 'extravagant' because it is a plain English word, more commonly used in everyday vocabulary. It is also more precise: the excessiveness language best captures the essence of a penalty. Clear and precise language aids comprehension and accessibility. Also, the use of plain English, precise language minimises the risk of misinterpretation.

Finally, the realigned comparative test adds the epithet 'manifestly'. It highlights that the excessive nature of the agreed damages amount must be 'Clear or obvious to the eye or mind';[38] or such that it would 'be at once apparent'.[39] Put differently, the clause's punitive

---

35 [1980] AC 827 (HL) 850. See similarly *O'Dea v Allstates Leasing System (WA) Pty Ltd* (1983) 152 CLR 359 (HCA) 369.

36 See *Cedar Meats (Aust) Pty Ltd v Five Star Lamb Pty Ltd* [2014] VSCA 32 [43], [51]-[52].

37 cf the use of the term 'disproportionate': see text following n 145 in ch 5.

38 Oxford Dictionaries <www.oxforddictionaries.com/definition/english/manifest> accessed 12 January 2015.

function must be blatant. Accordingly, the composite phrase 'manifest excess' emphasises that the test of validity applies constrictively, not expansively. That said, importantly the essential meaning of the phrase remains consistent with the current law's emphasis on substantial disproportionality, whereby only a 'very wide gulf'[40] suffices to render an agreed damages clause invalid.

II Context

Pursuant to the refined approach, an Anglo-Australian judge must decide whether the agreed damages amount is manifestly excessive when contrasted with the anticipated loss or gain, or the anticipated value of the infringed performance right. It has been held that this comparative test is 'relevant but no more than a guide' to determining whether a clause is penal.[41] Other factors, it is said, also come into play to decide whether the clause is valid, or not.[42] In fact, the comparative test alone robustly discriminates between valid and invalid agreed damages clauses. The key to understanding its ability to do so hinges on one crucial point: the test is applied *contextually*.

The comparative test must be applied contextually because corrective justice properly explains the Penalty Rule. Corrective justice calls for different approaches to different circumstances. That is, the corrective justice norm requires that the same principle be applied differently in particular cases because of a change in circumstances (in this context, a change in the degree of foreseeability and probability of an event occurring). The different treatment as between cases, however, is not evidence that the test of validity rests on several considerations, the refined comparative test being only one. Rather, it is the unfolding of a unified and coherent approach to the validity issue applied to different circumstances.

So, as refined, the test of validity encapsulates one comparative test, textured to context. Support for this view exists. For example, in *Zachariadis v Allforks Australia Pty Ltd*, the Victorian Court of Appeal held that the Extravagance Test is generally applicable, but it is

---

[39] *Clydebank Engineering and Shipbuilding Co Ltd v Don Jose Ramos Yzquierdo y Castaneda* [1905] AC 6 (HL) 10.

[40] *Alfred McAlpine Capital Projects Ltd v Tilebox Ltd* [2005] EWHC 281 (TCC), [2005] BLR 271 [48(4)].

[41] *Murray v Leisureplay plc* [2005] EWCA Civ 963, [2005] IRLR 946 [106(iv)].

[42] eg bargaining power: *M and J Polymers Ltd v Imerys Minerals Ltd* [2008] EWHC 344 (Com Ct), [2008] 1 Lloyd's Rep 541 [45]-[46]; *State of Tasmania v Leighton Contractors Pty Ltd* [2005] TASSC 133, (2005) 15 Tas R 243 [31(5)].

applied contextually, as the 'factual circumstances in which the contract is made will affect the application of that test'.[43]

Given its contextual nature, in practice the refined comparative test of validity must be applied according to particular principles. Four are noteworthy.

First, Anglo-Australian courts must consider the transaction's inherent and surrounding circumstances. As to the former, judges must pay close attention to the contract's nature and complexity,[44] for instance the nature of the contractual subject matter.[45] As to the latter, they must consider matters such as the 'historical background' of the parties' relationship,[46] and industry and market conditions.[47]

Second, the degree of exactitude the private law demands must vary in accordance with the ease or difficulty of the estimation. So, if the prospective loss, gain or value of the infringed right is very difficult to quantify, then a large discrepancy will be permissible. Considerable latitude is allowed in the relevant approximation.[48] As Chapter 9 argued, a 'generous' margin of appreciation is offered in most commercial cases, given that predicting likely future events is exceedingly difficult.[49] Conversely, if the anticipated loss, gain or value of the infringed right is relatively simple to quantify, then the private law requires a higher degree of exactitude. Much less latitude is permitted. This is perfectly consistent with the case law. Thus, Anglo-Australian courts have consistently held that where the anticipated loss is 'capable of exact definition',[50] the parties are granted a more narrow margin of appreciation.

This approach is not unknown to the private law. So, for a claimant to establish that a defendant is liable in negligence, she must show that the harm suffered was reasonably foreseeable. However, the degree to which the claimant must establish that the harm was foreseeable, rendering it *reasonably* foreseeable, varies depending on the circumstances. Contrast the following scenarios.[51] If the defendant's activity is socially valuable and it is very

---

[43] [2009] VSCA 258, (2009) 26 VR 47 [143]. See similarly *Esanda Finance Corp Ltd v Plessnig* (1989) 166 CLR 131 (HCA) 142; *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [75(ii)].

[44] *State of Tasmania v Leighton Contractors Pty Ltd* [2005] TASSC 133, (2005) 15 Tas R 243 [31(5)].

[45] *Esanda Finance Corp Ltd v Plessnig* (1989) 166 CLR 131 (HCA) 142 (second-hand goods).

[46] *Calcorp (Australia) Pty Ltd v 271 Collins Pty Ltd* [2010] VSCA 259, (2010) 29 VR 462 [21].

[47] Text following n 156 in ch 5.

[48] *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [75(iv)].

[49] *Cedar Meats (Aust) Pty Ltd v Five Star Lamb Pty Ltd* [2014] VSCA 32 [54].

[50] *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd* [1915] AC 79 (HL) 101; *Paciocco v Australia and New Zealand Banking Group Ltd* [2014] FCA 35, (2014) 309 ALR 249 [19(3)].

[51] *Overseas Tankship (UK) Ltd v Miller Steamship Co Pty Ltd (The Wagon Mound No 2)* [1967] 1 AC 617 (PC) 642-43.

difficult for the defendant to take precautions against a small risk of injury, the degree of probability with which the harm must be foreseeable will be high. Conversely, if the defendant's activity has no utility or is unlawful and the relevant risk is of serious injury, the degree of likelihood with which the loss must be foreseeable may be very low. Thus, a single test of reasonable foreseeability exists. However, courts apply that test contextually, taking into account factors such as the nature of the conduct, the degree of potential injury and the ability to minimise the risk. What courts require to satisfy the relevant standard of reasonable foresight, then, differs with the circumstances. The analogy with the contextual approach advocated here is obvious.

Third, in cases where approximation is very difficult, the nature of the inquiry is broad. However, it is important to be clear why this is so. It is said that in applying the test of validity in such cases, a judge must 'avoid nice calculations but … look at the question in the round' to render the test a 'broad and general one'.[52] This observation is misleading. A court does not eschew any rough calculation to ensure that the test of validity is kept broad. Rather, the test is inevitably broad because 'nice calculations' are simply not possible, given the difficulty in predetermination.

Finally, in cases where estimation is less difficult, the nature of the inquiry is more precise. This does not mean that the refined comparative test does not apply, or is not determinative, in cases where a reasonably accurate estimate is possible. Rather, it follows from the fact that the 'manifestly excessive' standard is applied contextually. Let us examine the point as regards agreed *compensatory* damages clauses. So, suppose that a reasonably precise estimate of the anticipated loss was possible at the contract date. If so, even a small discrepancy in percentage terms may lead to a finding of 'manifest excess'. On the one hand, even a small divergence in percentage terms may be 'excessive' because, as one could have straightforwardly predetermined the anticipated loss, the parties should have been able to include in their contract an agreed damages clause that closely reflected that loss. On the other hand, that excessiveness will be 'manifest' because, given the relative ease with which the future loss could have approximated, even a small difference in percentage terms will be immediately apparent to the judicial eye or mind.

Various examples might be given to buttress the point. One will have to suffice.[53] It is the 'central' case of *Zachariadis*. The parties entered into a contract for the hire of forklift

---

[52] *Murray v Leisureplay plc* [2005] EWCA Civ 963, [2005] IRLR 946 [106(xi)], [114].

[53] See also *O'Dea v Allstates Leasing System (WA) Pty Ltd* (1983) 152 CLR 359 (HCA) 369.

trucks. The contract contained an agreed damages clause providing that, upon termination, the hirer was to pay to the supplier all the charges that otherwise would have become payable from the contract's termination date to its expiry date. After terminating the agreement, the supplier sought to enforce the clause. The hirer contended it was penal. The Victorian Court of Appeal agreed.

Now, if the 'manifest excess' test applied bluntly, then it would invalidate only a clause that produced a 'substantial disproportion' in percentage terms between the agreed damages amount and the anticipated loss. However, on the facts, there seemingly did not exist a substantial divergence in percentage terms between these two figures. On this view, then, the clause was valid. It follows that either the outcome was wrong, or the test is not bluntly applied.

The latter is true. So, the case involved a chattel hire contract. In such agreements, it is possible to determine with some precision the expected loss upon termination. Consequently, the Court held that the parties would have had 'no difficulty' in producing an agreed damages sum that projected the anticipated loss relatively precisely.[54] As the agreed damages sum did not do so, it was penal. Put differently, although the disproportionality between the agreed damages sum and the anticipated loss may not have been great in percentage terms, in the circumstances a reasonably accurate estimate was possible. Given this, even the relatively small difference in percentage terms sufficed to render the amount quite obviously—thus 'manifestly'—excessive.

In sum, the refined comparative test is the sole test. However, it applies contextually. If, evaluated contextually, the agreed damages sum is not excessive, then it is valid.

III The comparative standard

Under the refined test of validity, what is the appropriate standard of comparison to apply? For example, if a court is assessing an agreed disgorgement damages clause, does it compare the agreed disgorgement damages sum with B's likely gain; B's possible gain based on realistic and not far-fetched events; or the greatest possible gain that B might make?

One must begin, as always, with the rule's true justification. Two features of the corrective justice norm are important here. For one, Chapter 8 established that, to be valid, an

---

[54] *Zachariadis v Allforks Australia Pty Ltd* [2009] VSCA 258, (2009) 26 VR 47 [157]. The Court of Appeal expressly distinguished the case of *Yarra Capital Group Pty Ltd v Sklash Pty Ltd* [2006] VSCA 109, where predicting the anticipated loss was very difficult: see [137], [157].

agreed damages clause must instate the next-next best position. To do so, an agreed damages sum must notionally represent a legitimate response that would come to pass if future events unfolded in a probable, or possible but not far-fetched, manner. For another, as Chapter 8 noted, pursuant to its fairness principle corrective justice treats the parties as equals. This requires the private law to apply standards that mediate between the parties' interests.[55] If the law privileges one party's position, then it fails to treat the parties fairly, and thus fails to do justice as between the parties.

The first correctivist tenet grounds the following proposition: the refined comparative test requires that a judge compare the agreed damages amount with the loss, gain or value of the infringed right that would arise if future events unfolded in a probable, or possible but not far-fetched, way. The second correctivist tenet leads to the view that a court must disregard future events that are possible but far-fetched, or improbable. These statements require unpacking. Three points arise.

To start, it is settled law that the test of validity takes account of probable or likely consequences. On several occasions, the Australian High Court has held that an agreed damages sum will be penal 'if it is out of all proportion to damage *likely* to be suffered as a result of breach'.[56] In many cases, the English Court of Appeal has held similarly.[57]

Next, it has long been suggested that the benchmark for the comparative test includes events that may possibly occur, but whose possible occurrence is not far-fetched in nature. So, in describing the appropriate comparative standard, various Anglo-Australian appellate courts have used the language of probability *and possibility*.[58]

However, the relevant possibilities have been limited. Take, for instance, *Esanda Finance Corp Ltd v Plessnig*.[59] The parties entered into a hire purchase contract concerning a second hand prime mover. The contract provided that, upon termination, the owners were entitled to repossess the prime mover and sell it. The agreement contained an agreed damages clause. Relevantly, that clause did not require the owner to provide the hirers with a refund of any surplus resulting from the owner selling the repossessed trucks at a profit. The hirers

55 Allan Beever, *Rediscovering the Law of Negligence* (Hart Publishing 2007) 84.

56 *AMEV-UDC Finance Ltd v Austin* (1986) 162 CLR 170 (HCA) 190 (emphasis added); *Ringrow Pty Ltd v BP Australia Pty Ltd* [2005] HCA 71, (2005) 224 CLR 656 [27], [31] (emphasis added).

57 *Robophone Facilities Ltd v Blank* [1966] 1 WLR 1428 (CA) 1446; *Makdessi v Cavendish Square Holdings BV* [2013] EWCA Civ 1539, [2014] 2 All ER (Comm) 125 [79], [81].

58 *Imperial Tobacco Co (of Great Britain and Ireland) Ltd v Parslay* [1936] 2 All ER 515 (CA) 521; *Lamson Store Service Co Ltd v Russell Wilkins and Sons Ltd* (1906) 4 CLR 672 (HCA) 681-82, 689.

59 (1989) 166 CLR 131 (HCA).

argued that the omission meant the clause was penal. The Australian High Court disagreed. Wilson and Toohey JJ held that the 'mere possibility … of an excess' was insufficient to render the clause penal.[60] This was because the argument was based on the assumption that the vehicle would retain a significant resale value. That assumption was far-fetched, given the vehicle was second hand goods *and* it was used to engage in heavy haulage.

The refined comparative test takes this stance because corrective justice requires it. Corrective justice demands that the private law be fair. As such, the refined test of validity does not, and cannot, take account of *unlikely* events or consequences, as this would privilege B's position. As the Board noted in *Philips Hong Kong Ltd v A-G of Hong Kong*, it would allow B easily to impugn an agreed damages clause by postulating an unlikely, or possible but far-fetched, scenario that rendered the agreed damages sum manifestly excessive.[61]

Lastly, the suggested approach is inconsistent with Lord Dunedin's precise formulation of the Extravagance Test in *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd*.[62] His Lordship held that an agreed compensatory damages sum is penal if it is extravagant when compared with 'the *greatest* loss that could conceivably be proved to have followed from the breach'.[63] If Lord Dunedin is rightly interpreted as saying that an agreed compensatory damages clause is valid even if manifestly excessive as compared with the greatest possible loss that may conceivably occur, however improbable, then his Lordship was mistaken. If the correct baseline were the greatest possible anticipated loss, then the relevant standard of discrepancy would not be one of 'extravagance' or 'manifest excess'.

Thus, if the test of validity rightly compared the agreed damages amount with the greatest possible loss, then the baseline for comparison would be *the highest estimate or the top end of any range of estimates*. Accordingly, the permissible discrepancy between this figure and the agreed damages amount would have to be relatively small. If the discrepancy were large, then a judge could be confident enough (on the balance of probabilities) that at least part of the agreed damages sum would end up operating punitively, and the clause would be invalid. In other words, if the benchmark for comparison is the greatest possible loss, then the test of validity must afford the parties only a small margin of appreciation, not a wide one. But, as his Lordship recognised, this is not the case.

---

[60] *Esanda Finance Corp Ltd v Plessnig* (1989) 166 CLR 131 (HCA) 142. See also 152 (Brennan J).

[61] (1993) 61 BLR 41 (PC) 59.

[62] [1915] AC 79 (HL).

[63] *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd* [1915] AC 79 (HL) 87 (emphasis added). See also *GWC Property Group Pty Ltd v Higginson* [2014] QSC 264 [39].

D Applied

This section explores three famous cases, chosen because they involve a very similar fact pattern. It demonstrates that the equilibrial rule would reach the same result as the court did in each case, but do so through more elegant and powerful reasoning: in each case, relying on the legitimate function of substitution. In short, this section shows that the proposed approach is perspicuous, and the case law reflects it.

The first case is the most 'central' case of all: *Dunlop*. Dunlop manufactured motor tyres, covers and tubes. It supplied these to its dealer, New Garage, on terms. New Garage agreed not to do any of the following: tamper with Dunlop's markings on the covers or tubes; sell or offer the goods to customers at prices below Dunlop's current list prices; sell the goods to persons whose supplies Dunlop had decided to suspend; or exhibit or export the goods without Dunlop's consent. Under clause 5, New Garage agreed to pay to Dunlop £5 as 'liquidated damages' for each product sold or offered in breach of the agreement. New Garage sold covers and tubes at prices below Dunlop's set prices. Dunlop sued. In its defence, New Garage argued that clause 5 was penal. The House of Lords unanimously held the clause was valid.

Their Lordships adopted a traditional, loss-based analysis. Given there were several ways New Garage might relevantly breach the agreement, the House of Lords held that the term failed the Indiscriminate Cause Test. This raised a presumption of penalty. However, the loss likely to result from any breach was 'of such an uncertain nature'[64] that it could not be accurately assessed. It followed that the clause passed the Impossibility Test, which rebutted the presumption. One aspect of the anticipated loss that was apparently impossible to assess was the undermining of Dunlop's agency system of marketing and distribution. In Lord Dunedin's view, the undercutting of the integrity of Dunlop's wholesale distribution system constituted an 'indirect' loss which was 'impossible to forecast'.[65] As Chapter 5 illustrated, however, a loss-based interpretation of the Impossibility Test renders that test hopelessly flawed.

Compare how the refined test of validity applies on the facts. So, was clause 5 manifestly excessive in light of the provision's legitimate function? Two questions must be asked.

---

[64] *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd* [1915] AC 79 (HL) 96.

[65] *Dunlop Pneumatic Tyre Co Ltd v New Garage and Motor Co Ltd* [1915] AC 79 (HL) 88.

The first question is this: what was the legitimate function that the agreed damages sum performed? As Chen-Wishart argues, 'Dunlop did not show that any breach would cause any quantifiable loss at all.'[66] In truth, the anticipated 'loss' was impossible to estimate because it did not in fact exist. Rather, according to the Australian High Court in *Andrews*, in *Dunlop* 'the critical issue … was whether the sum agreed was commensurate with the interest protected by the bargain'.[67] As Chapter 5 demonstrated, the 'interest protected by the bargain' is best interpreted as referring to Dunlop's performance right.[68] Thus, correctly conceptualised, the clause concerned substituting for the value of the infringed performance right, not compensating for loss. The proper legitimate function, then, was substitution.

The second question may be put thus: applying the refined comparative test, did the clause impose upon New Garage a detriment that was manifestly excessive when compared with the anticipated value of Dunlop's infringed performance right? Here, the value of the right is quantified adopting a reasonable, hypothetical bargain approach. So, we assume a hypothetical bargain occurred between a reasonable, willing car accessory manufacturer and a reasonable, willing dealer. The question posed, on the contract date, was this: if such a bargain occurred in the future, would the manufacturer release its right to performance, and would the dealer buy that release, for an amount somewhere generally in the vicinity of £5 for each product sold? The answer is 'yes'. It follows that the agreed substitutive damages amount under clause 5 was not obviously excessive. The clause was therefore valid.

The second case is *Imperial Tobacco Co (of Great Britain and Ireland) Ltd v Parslay*, another 'central' case.[69] Imperial manufactured tobacco products. It supplied these to a retail tobacconist, Parslay. The parties entered into a resale price maintenance agreement. Parslay agreed not sell any tobacco product to any customer at a price lower than Imperial's then fixed price. Further, pursuant to an agreed damages clause, Parslay agreed to pay Imperial £15 for each tobacco product he sold in breach. Parslay committed several breaches; Imperial sought to enforce the clause. Parslay claimed it was penal. The English Court of Appeal unanimously held that the clause was a liquidated damages clause.

Their Lordships held that the case was one 'in principle like that of the *Dunlop*', such that the reasoning underpinning the House of Lords's conclusion 'equally impelled' the Court

---

[66] Mindy Chen-Wishart, 'Controlling the Power to Agree Damages' in Peter Birks (ed), *Wrongs and Remedies in the Twenty-First Century* (Clarendon Press 1996) 287 (footnote omitted).

[67] *Andrews v Australia and New Zealand Banking Group Ltd*, in *Dunlop* [2012] HCA 30, (2012) 247 CLR 205 [75].

[68] Text accompanying and following n 89 in ch 5.

[69] [1936] 2 All ER 515 (CA).

of Appeal's decision on the facts.[70] Thus, the Court held that the loss that would arise from the undercutting of Imperial's list prices, and the disturbance to its marketing and distribution system, was 'impossible to specify in advance'.[71] It followed that the agreed damages clause satisfied the Impossibility Test, and was valid.

Contrast the reasoning if one applies the refined test of validity. Clearly, for the reasons explained regarding *Dunlop*, the legitimate function in play is substitution. The main question is whether, applying the refined comparative test, the agreed substitutive damages clause imposed a detriment that was manifestly excessive when compared with the anticipated value of Imperial's infringed performance right.

Again, adopting a reasonable, hypothetical bargain approach, one must ask the following question as at the contract date: in a future hypothetical bargain, would a reasonable tobacco manufacturer agree to release its performance right, and would a reasonable retailer buy that release, for an amount somewhere generally in the vicinity of £15 for each product sold? The answer is 'possibly'. So, two reasonable parties might enter into such a hypothetical release agreement for their own commercial reasons: on the retailer's side, if she wished to use a particular tobacco product as a loss leader; on the manufacturer's side, if the dealer making such a request was sufficiently geographically isolated. Although one may say the amount was excessive, it was not *obviously* excessive. The clause, therefore, was properly a liquidated substitutive damages clause.

The final case is the 'standard' case of *Ford Motor Co (England) Ltd v Armstrong*.[72] Ford was a car manufacturer. It agreed to sell cars to Armstrong, a car dealer, for on sale. In return, Armstrong agreed not to resell any car or any car part at a price under a listed price; not to resell the goods to other dealers; and not to exhibit a car without Ford's permission. Further, Armstrong agreed to pay Ford £250 in 'agreed damages' for any breach.[73] Armstrong sold five cars at a price below that listed by Ford. A dispute ensued as to the term's validity. A majority of the English Court of Appeal held the clause was penal.

Their Lordships all purported to apply a traditional, loss-based analysis. However, the Court was split on the outcome that followed from it. Buckley LJ, with whom Pickford LJ agreed, held that the case involved the same prospective 'indirect' loss as existed in *Dunlop*.

---

[70] *Imperial Tobacco Co (of Great Britain and Ireland) Ltd v Parslay* [1936] 2 All ER 515 (CA) 524, 525.

[71] *Imperial Tobacco Co (of Great Britain and Ireland) Ltd v Parslay* [1936] 2 All ER 515 (CA) 521. See also 524, 525.

[72] (1915) 31 TLR 267 (CA).

[73] *Ford Motor Co (England) Ltd v Armstrong* (1915) 31 TLR 267 (CA) 268.

His Lordship held that, although the clause failed the Indiscriminate Cause Test, the anticipated loss was 'incapable of being precisely ascertained',[74] and thus it passed the Impossibility Test. However, Buckley LJ then distinguished the facts from those in *Dunlop* on the basis that here the agreed damages sum was substantially higher. Accordingly, the clause was penal. Conversely, Bankes LJ dissented. His Lordship held that, applying the *Dunlop* tests, the clause was valid.

Putting aside the intractable difficulties plaguing the loss-based view of the Impossibility Test, Bankes LJ's view has some merit. Under the Impossibility Test, the fact that pre-estimation is essentially impossible will not render a clause penal; rather, it is *more likely* that the clause will be upheld. The Court was clear in its view that the anticipated loss was impossible to pre-assess. If so, then presumably on a loss-based view the clause was properly valid, and the fact the agreed damages amount was substantially higher than that in *Dunlop* was immaterial.

The refined test of validity reaches the same conclusion, without the controversy. The provision properly served a substitutive function. Thus, the only question is whether the result is explicable in terms of manifest excess. It is. No reasonable, willing hypothetical dealer would pay an anticipated release fee of £250. The sum was *clearly* excessive. The clause was rightly penal.

In conclusion, when conducting its review of the Penalty Rule, the Law Commission formed the 'clear view' that the test of validity had, 'on the whole, worked satisfactorily'.[75] That statement is mostly correct. This section has outlined the interpretive steps required to ensure the test works satisfactorily, full stop.

4 Consequences

This final section analyses consequential matters. It contains two parts. In the first part, the section examines the consequences of a court upholding an agreed damages clause as a liquidated damages clause under the refined test of validity. In the second part, it explicates what is to occur when, after applying the improved test, a court invalidates the clause and holds that it is a penalty.

[74] *Ford Motor Co (England) Ltd v Armstrong* (1915) 31 TLR 267 (CA) 268.

[75] Law Commission, *Penalty Clauses and Forfeiture of Monies Paid* (Law Com WP No 61, 1975) [41].

A Liquidated damages clauses

Chapter 6 examined the consequences of a court validating an agreed damages clause. It concluded that liquidated damages are generally in the nature of a debt, not damages. Accordingly, A may recover from B the liquidated damages sum in the same way as she would recover any other debt. Further, in restating the law on consequences, Chapter 6 contrasted two situations concerning liquidated compensatory damages clauses.

On the one hand, where A's incurred loss is *less* than the liquidated compensatory damages amount, she may enforce the clause nonetheless.

On the other hand, where A's incurred loss is *more* than the liquidated compensatory damages amount, the liquidated compensatory damages clause provides an exclusive remedy for compensatory damages for B's default. However, the exclusivity is limited. It only applies for the particular breach or breaches of obligation, and kinds of consequential loss, falling within the scope of the liquidated compensatory damages clause.

The refined test of validity holds that an agreed damages clause is valid if the detriment it imposes is not manifestly excessive when compared with the amount of A's expected loss or gain, or the expected value of A's infringed performance right. Does that change matters? As regards liquidated compensatory damages clauses, the answer is that it does not. The consequences of validity will remain the same. However, given the preventing punishment principle that underlies the rule, Anglo-Australian private law also rightly enforces liquidated damages that are gain-based or substitutive in nature. Hence, the principles stated in relation to liquidated compensatory damages clauses must apply *mutatis mutandis* to liquidated substitutive damages clauses, liquidated restitutionary damages clauses and liquidated disgorgement damages clauses. For instance, if B's proved gain is *less* than the liquidated disgorgement damages amount, she may enforce the clause nonetheless. Equally, if the proven value of A's infringed performance right is *more* than the liquidated substitutive damages amount, the liquidated damages clause provides an exhaustive remedy. Consequently, A cannot recover from B general substitutive damages instead of, or in addition to, the liquidated substitutive damages sum.[76]

B Penalty clauses

[76] *Diestal v Stevenson* [1906] 2 KB 345 (KB) 351.

Chapter 6 also assessed the consequences of a court impugning an agreed damages clause under the Penalty Rule. It established that, correctly understood, penalties are generally akin to a debt; and partially unenforceable, not void.

In its restatement of the law on the consequential relief, Chapter 6 considered the case where a court invalidates an agreed compensatory damages clause. It examined two situations. In each context, the chapter answered two questions. May A elect between two means of recovery, choosing whether to institute against B a general damages claim or sue B on the penalty clause? Also, what is nature of the proper means of recovery?

The first situation was where A's incurred loss is *less* than the penalty amount. As to the election question, although it appears that A has a right to elect, in truth no such right exists. As to the means question, Anglo-Australian private law adopts the compression approach. That is, as the penalty clause is partially invalid, the clause is 'compressed' to the extent to which it is enforceable, which means B's liability is 'cut down' to an amount equivalent to A's incurred loss. A is then entitled to bring a claim partially to enforce the penalty clause. Further, if B has already satisfied his obligation to pay A under the penalty clause, B will obtain an unjust enrichment claim right to restitution from A. He will recover a sum reflecting the unenforceable, excessive component of the penalty clause, which is the difference in amount between the penal sum and A's incurred loss.

The second situation was where A's incurred loss *exceeds* the penalty amount. On the election issue, any right to elect is an illusion. As best understood, Anglo-Australian courts *only* partially enforce the clause. On the means issue, A has the power to sue on the penalty clause, enforcing it up to a sum equal to the amount of her incurred loss. As A's loss exceeds the penal sum, no compression is needed. A may sue for the entire penal amount. Logically, then, she cannot recover more than the penal sum.

Given the improved test of validity, must the law refine its approach to the consequences of invalidity? The answer is that Anglo-Australian private law must apply the restated regime analogically to penalty clauses partially involving legitimate, non-compensatory responses. So, a penalty clause that partially performs a legitimate, substitutive or gain-based function is partially unenforceable. The clause is 'compressed' to the extent to which it is enforceable. In other words, the penal element is suppressed. A retains a right of action against B based on the penalty clause. She may partially enforce the clause to the extent of its validity: that is, to recover an amount equivalent to B's proved gain or the proven value of A's infringed performance right. It follows that the principles stated in both situations above apply by analogy to gain-based penalty clauses, and substitutive penalty clauses. For example, if a court invalidates an agreed disgorgement damages clause and B's gain is *less* than the penalty amount, in principle A may exercise her right of action to recover a sum equal to the amount of B's proved gain. Similarly, if a court impugns an agreed substitutive damages clause and the proven value of A's infringed performance right is *more* than the penal sum, A can recover the penal sum only.

There is one final point. In the 'core case', a penalty clause imposes upon B a duty to pay a sum of money to A. Accordingly, how a partial enforcement regime works is clear: A will have a right of action partially to enforce the clause up to, at a maximum, the penalty amount. However, this thesis has demonstrated that a clause requiring B to perform a non-monetary obligation may also be penal. In these circumstances, how will a partial enforcement regime generally[77] operate? Like this.[78] A court will order that B must pay a sum that compensates A for her incurred loss, and that upon doing so B's penal obligation is discharged. This solution effectively implements in the circumstances the 'scaling down' exercise that Anglo-Australian courts have supported. In short, this general approach represents 'the equivalent in the different circumstances … to the automatic scaling down of a (pecuniary) penalty clause'.[79]

5 Conclusion

This chapter has articulated a refined Penalty Rule that achieves equilibrium between the penalty doctrine as restated and its theoretical justification. The equilibrial rule thus casts the doctrine in the best possible light, all things considered. The chapter explained how the refined rule operates at each of the three stages of inquiry, as follows.

First, the chapter worked by analogy from the 'core case' to develop a refined jurisdictional test. In doing so, it demonstrated that at the heart of the concept of a substantive breach of contract lies a false illusion of choice. The chapter then applied the realigned test to a series of difficult examples, illustrating that the improved rule achieves a tight fit, but does so through elegant and unified reasoning.

Second, the chapter framed and elucidated a refined test of validity. It showed that at the core of the refined comparative test lies the idea of 'manifest excess'. The chapter

[77] cf *Jobson v Johnson* [1989] 1 WLR 1026 (CA) 1037, 1045.

[78] For simplicity, the following will only deal with the 'core case' of a compensatory penalty clause.

[79] *Jobson v Johnson* [1989] 1 WLR 1026 (CA) 1045.

explained that the comparative test is applied contextually, but requires a particular comparative standard. It then systematically applied the improved test of validity to three significant cases involving similar facts. In each case, the refined rule produced the same outcome as the relevant court, but the reasoned path to that result was more elegant and powerful.

Finally, the chapter examined consequential issues. It showed that whether a court validates or invalidates an agreed *compensatory* damages clause under the refined test of validity, the consequences will remain the same. Conversely, the chapter illustrated that whether a judge upholds or impugns an agreed *gain-based* damages clause, or an agreed *substitutive* damages clause, only small interpretive steps are needed properly to refine the court's approach to the consequences of validity or invalidity respectively.

One final point exists. It concerns the fourth and fifth aims of this thesis, set out in Chapter 1.

The fourth aim was to refine the Penalty Rule with its proper theoretical justification in mind. That aim has now been accomplished. Through the various examples considered, this chapter has shown that the refined rule is capable of better explaining the seminal past decisions; can provide a sound guide for judicial decision-making; and is able to be clearly applied in commercial cases.

The final aim was self-reflectively to evaluate the refined rule. This chapter demonstrated that the improved rule paints the best picture of the law that can be designed. The chapter showed that the refined rule is normatively appealing, as all aspects of it follow from corrective justice; and that the improved rule is theoretically defensible, as an ultimate appellate court judge could fashion it by taking only small interpretive steps. Further, the applications of the refined rule to various examples show that it is able successfully to function in commercial practice. Consequently, this final aim of this thesis has also been achieved.

11 CONCLUSION

This thesis has adopted a rights-based, interpretive and formalist methodology to examine the Penalty Rule in Anglo-Australian commercial law. It has sought to restate the law, theoretically justify it, and then refine the restated law in light of its proper theoretical justification. This process aimed to produce a refined rule that achieves equilibrium between the doctrine as it is found, and the best explanation of that doctrine. In other words, the thesis has attempted to cast the Penalty Rule in the best possible light, all things considered.

The refined Penalty Rule can be succinctly stated. It demands three stages of analytical inquiry. First, an Anglo-Australian court must take jurisdiction. A court has jurisdiction if the clause imposes, in A's favour, a *detriment* upon B as a result of B's *default*. A detriment is an obligation to pay money, transfer a right or extinguish a right. A default is a consent-based actual breach, or a consent-based substantive breach. Second, a judge must apply the test of validity, as at the contract date. An agreed damages clause is invalid if it requires B to incur a detriment that is *manifestly excessive* in light of the clause's *legitimate function*. Legitimate functions include substitution, compensation, restitution and disgorgement. To determine if the clause is manifestly excessive, a court must compare the agreed damages sum or value with the amount of A's anticipated loss or gain, or the anticipated value of A's infringed performance right. Finally, the court must determine the consequences of validity, or invalidity. A has a claim right to enforce a liquidated damages clause against B according to its terms. Conversely, a penalty clause is partially unenforceable. A has a right of action to enforce the clause against B only to the extent of its validity. Its validity will extend to an amount equivalent to A's incurred loss, B's proved gain, or the proven value of A's infringed performance right. In summary, that is the whole of the refined Penalty Rule. It is the best picture of the rule that can be rendered.

Given the expansive range of legitimate functions, under the refined rule it is very likely that an agreed damages clause will be upheld. However, that is very much in line with the current law. The refined approach, then, will affect not so much what Anglo-Australian private law does in this area, but how we understand what it does.

The Penalty Rule is not concerned with avoiding substantive unconscionability, promoting efficiency or preventing 'overcompensation'. Rightly understood, a two-level theory justifies the Penalty Rule.

At the lower level, the classificatory theory of the rule holds that the right to a penalty is a consensual punishment. The Penalty Rule deems that right to be illegitimate, and partially unenforceable. Accordingly, the rule encapsulates a principle of preventing consensual civil punishment. Two important consequences follow. For one, the Penalty Rule is proscriptive, and not prescriptive. As best interpreted, the doctrine prohibits any agreed damages clause that operates punitively. It does not demand that all agreed damages clauses must be compensatory. For another, penalty clauses are a punitive damages analogue. Anglo-Australian private law prohibits a party enforcing the former, and generally prevents a judge awarding the latter, based on a similar principle of prohibiting punishment.

At the higher level, a corrective justice theory of the doctrine holds that the principle against consensual punishment enunciates what corrective justice demands in the agreed damages clauses context. Consequently, the Penalty Rule is a manifestation of corrective justice. As such, the corrective justice norm provides the rule's ultimate theoretical explanation. Corrective justice supplies principled reasons establishing why punishment is properly an unavailable response to a general law wrong, and hence why punitive damages are unacceptable in the private law. These reasons apply with equal force to render punishment illegitimate where it responds to consent, and thus they justify the Penalty Rule. Three common reasons exist: punishment is unfair; it fails to correct satisfactorily the relevant injustice; and it stands outside the structure of civil liability. However, there exists a further reason applying solely to penalty clauses that renders them illegitimate in private law, whatever the case may be in the law of wrongs: namely, that consensual punishment is not generally deterrent in effect. Thus, when it comes to private law's need to prevent punishment, consensual punishment is the a fortiori case.

The classificatory, correctivist theory of the Penalty Rule is robust. On four criteria for evaluating private law theories, this dual theory scores extremely well. First, this explanation generates outcomes that are consistent with the decisions made in the cases. In other words, it displays a tight fit. It also possesses great explanatory power. Second, this rights-based theory strictly conforms to the courts' general reasoning, and is thus highly transparent. Third, this rationalisation renders the private law weakly coherent, for it reveals the law to be internally consistent; and also satisfies a more demanding, strong notion of coherence, as it contributes towards the private law, or a part of it, potentially being understood as a conceptually unified system. Finally, the two-level theory is practically beneficial, as the theory permits a practitioner to explain the rule in a way that will allow a commercial client easily and efficiently to understand and apply it.

Writing in 1961, Wedderburn surmised that the Penalty Rule was 'uncertain, inelegant and unjust'.[1] The rule, however, did not deserve these criticisms then. Nor does it merit them today. In fact, when viewed in the best possible light, the rule is certain, elegant and (correctively) just. Nothing more is required of a private law doctrine. But, of course, nothing less will suffice.

[1] KW Wedderburn, 'Hire-Purchase—Penalties—Freedom of Contract' [1961] CLJ 156, 157.

BIBLIOGRAPHY

1 Books

— — *Carter on Contract* (service 37, April 2014)

Andrews N, *Contract Law* (CUP 2011)

Atiyah PS, *Essays on Contract* (Clarendon Press 1990)

Beale H, *Remedies for Breach of Contract* (Sweet and Maxwell 1980)

— — (gen ed), *Chitty on Contracts* (2nd supp, 31st edn, Sweet and Maxwell 2013)

Beatson J, Burrows A and Cartwright J, *Anson's Law of Contract* (29th edn, OUP 2010)

Beever A, *Rediscovering the Law of Negligence* (Hart Publishing 2007)

— — 'Justice and Punishment in Tort: A Comparative Theoretical Analysis' in Rickett C (ed), *Justifying Private Law Remedies* (Hart Publishing 2008)

— — *The Law of Private Nuisance* (Hart Publishing 2013)

Birks P, *An Introduction to the Law of Restitution* (Clarendon Press 1985)

— — 'Definition and Division: A Meditation on Institutes 3.13' in Birks P (ed), *The Classification of Obligations* (Clarendon Press 1997)

— — *Unjust Enrichment* (2nd edn, OUP 2005)

Blackstone, *Commentaries on the Law of England*

Bridge M, *The Sale of Goods* (3rd edn, OUP 2014)

Bryan M, 'Equitable Relief from Forfeiture: Performance or Restitution?' in Rickett C (ed), *Justifying Private Law Remedies* (Hart Publishing 2008)

Burrows A, *Remedies for Torts and Breach of Contract* (2nd edn, Butterworths 1994)

— — *Remedies for Torts and Breach of Contract* (3rd edn, OUP 2004)

— — 'The Relationship between Property and Unjust Enrichment: Some Unresolved Issues' in Degeling S and Edelman J (eds), *Unjust Enrichment in Commercial Law* (Lawbook Company 2008)

— — *The Law of Restitution* (3rd edn, OUP 2011)

— — (ed), *English Private Law* (3rd edn, OUP 2013)

Carter JW, *Carter's Breach of Contract* (LexisNexis Butterworths 2011)

Chen-Wishart M, 'Controlling the Power to Agree Damages' in Birks P (ed), *Wrongs and Remedies in the Twenty-First Century* (Clarendon Press 1996)

— — *Contract Law* (4th edn, OUP 2012)

Collins H, 'Fairness in Agreed Remedies' in Willett C (ed), *Aspects of Fairness in Contract* (Blackstone 1996)

— — *The Law of Contract* (4th edn, LexisNexis 2003)

Dal Pont GE, *Equity and Trusts in Australia* (5th edn, Lawbook Company 2011)

Downes T, 'Rethinking Penalty Clauses' in Birks P (ed), *Wrongs and Remedies in the Twenty-First Century* (Clarendon Press 1996)

Edelman J, 'Gain-Based Damages and Compensation' in Burrows A and Rodger A (eds), *Mapping the Law: Essays in Memory of Birks P* (OUP 2006)

— — 'In Defence of Exemplary Damages' in Rickett C (ed), *Justifying Private Law Remedies* (Hart Publishing 2008)

Fried C, *Contract As Promise: A Theory of Contractual Obligation* (Harvard UP 1981)

Furmston M (gen ed), *The Law of Contract* (4th edn, Butterworths 2010)

Furst S and Ramsey V, *Keating on Construction Contracts* (8th edn, Sweet and Maxwell 2006)

Gardner J, 'Introduction' in Hart HLA, *Punishment and Responsibility: Essays in the Philosophy of Law* (Gardner J ed, 2nd edn, OUP 2008)

Greig DW and Davis JLR, *The Law of Contract* (Lawbook Company 1987)

Gullifer L, 'Agreed Remedies' in Burrows A and Peel E (eds), *Commercial Remedies: Current Issues and Problems* (OUP 2003)

Harris D, Campbell D and Halson R, *Remedies in Contract and Tort* (2nd edn, Butterworths 2002)

Hart HLA, *The Concept of Law* (3rd edn, OUP 2012)

Heydon JD and Leeming MJ, *Cases and Materials on Equity and Trusts* (8th edn, LexisNexis Butterworths 2011)

Holmes OW, *The Common Law* (Little Brown 1881)

Ibbetson DJ, *A Historical Introduction to the Law of Obligations* (OUP 2001)

Mason K, Carter JW and Tolhurst G, *Mason and Carter's Restitution Law in Australia* (2nd edn, LexisNexis Butterworths 2008)

McBride MJ, 'Rights and the Basis of Tort Law' in Nolan D and Robertson A (eds), *Rights and Private Law* (Hart Publishing 2012)

McGhee J (gen ed), *Snell's Equity* (33rd edn, Sweet and Maxwell 2014)

McGregor H, *McGregor on Damages* (19th edn, Sweet and Maxwell 2014)

McInnes M, 'Taxonomic Lessons for the Supreme Court of Canada' in Rickett C and Grantham R (eds), *Structure and Justification in Private Law* (Hart Publishing 2008)

McKendrick E, 'Liquidated Damages and Related Clauses in Claims involving Chattels' in Palmer N and McKendrick E (eds), *Interests in Goods* (2nd edn, LLP 1998)

— — *Contract Law* (10th edn, Palgrave Macmillan 2013)

Nolan D and Robertson A, 'Rights and Private Law' in Nolan D and Robertson A (eds), *Rights and Private Law* (Hart Publishing 2012)

Peel E, *The Law of Contract* (13th edn, Sweet and Maxwell 2011)

Posner RA, *Economic Analysis of Law* (8th edn, Aspen Publishers 2011)

Rickett C, 'Unconscionability in Commercial Law' in Lowry J and Mistelis L (eds), *Commercial Law: Perspectives and Practice* (Lexis Nexis Butterworths 2006)

Rossiter C, *Penalties and Forfeiture* (Lawbook Company 1992)

— — 'Relief Against Penalties' in Parkinson P (ed), *Principles of Equity* (2nd edn, Lawbook Company 2002)

Seddon N, Bigwood R and Ellinghaus M, *Cheshire and Fifoot Law of Contract* (10th Aus edn, LexisNexis Butterworths 2012)

Shelton J, 'Agreed Remedies: Comment' in Burrows A and Peel E (eds), *Commercial Remedies: Current Issues and Problems* (OUP 2003)

Smith SA, *Contract Theory* (OUP 2004)

— — *Atiyah's Introduction to the Law of Contract* (6th edn, Clarendon Press 2005)

Stevens R, *Torts and Rights* (OUP 2007)

— — 'Damages and the Right to Performance: a Golden Victory or Not?' in Neyers JW, Bronaugh R and Pitel SGA (eds), *Exploring Contract Law* (Hart Publishing 2009)

— — 'Rights and Other Things' in Nolan D and Robertson A (eds), *Rights and Private Law* (Hart Publishing 2012)

— — 'Private Rights and Public Wrongs' in Dyson M (ed), *Unravelling Tort and Crime* (CUP 2014)

Stone M, 'Formalism' in Coleman J and Shapiro S (eds), *The Oxford Handbook of Jurisprudence and Philosophy of Law* (OUP 2002)

Tettenborn A (gen ed), *The Law of Damages* (2nd edn, LexisNexis 2010)

Tilbury M, 'Remedies and the Classification of Obligations' in Robertson A (ed), *The Law of Obligations: Connections and Boundaries* (UCL 2004)

Virgo G, *The Principles of the Law of Restitution* (2nd edn, OUP 2006)

Waddams SM, *The Law of Contracts* (6th edn, Canada Law Book 2010)

Weinrib EJ, *The Idea of Private Law* (Harvard UP 1995)

— — 'Two Conceptions of Remedies' in Rickett C (ed), *Justifying Private Law Remedies* (Hart Publishing 2008)

— — *Corrective Justice* (OUP 2012)

Worthington S, *Equity* (2nd edn, OUP 2006)

— — 'What is Left of Equity's Relief Against Forfeiture?' in Bant E and Harding M (eds), *Exploring Private Law* (CUP 2010)

Zakrzewski R, *Remedies Reclassified* (OUP 2005)

2 Articles

Barton JL, 'Penalties and Damages' (1976) 92 LQR 20

Beever A, 'The Structure of Aggravated and Exemplary Damages' (2003) 23 OJLS 87

— — and Rickett C, 'Interpretive Legal Theory and the Academic Lawyer' (2005) 68 MLR 320

Birks P, 'Equity in the Modern Law: An Exercise in Taxonomy' (1996) 26 UWAL

— — 'The Law of Restitution at the End of an Epoch' (1999) 28 UWAL

— — 'The Law of Unjust Enrichment: A Millennial Resolution' (1999) Sing JLS 318

— — 'Recovering Value Transferred under an Illegal Contract' (2000) 1 Theo Inq L 155

— — 'Rights, Wrongs, and Remedies' (2000) 20 OJLS 1

Carter JW and Courtney W, 'Indemnities against Breach of Contract as Agreed Damages Clauses' [2012] JBL 555

— — and others, 'Contractual Penalties: Resurrecting the Equitable Jurisdiction' (2013) 30 JCL

— — and Peden E, 'A Good Faith Perspective on Liquidated Damages' (2007) 23 JCL 157

— — and Tilbury M, 'Remedial Choice and Contract Drafting' (1998) 13 JCL 5

Clarke MJR, 'Penalties, Forfeiture and Dilution' [1989] AMPLA Ybk 1

Clarkson KW, Miller RL and Muris TJ, 'Liquidated Damages v Penalties: Sense or Nonsense?' [1978] Wis L Rev 351

Coase RH, 'The Problem of Social Cost' (1960) 3 JLE 1

Conte C, 'The Jurisdiction to Relieve against Penalties and Forfeitures: Time for a Rethink' (2010) 126 LQR 529

— — 'From Only the 'Bottom-up'? Legitimate Forms of Judicial Reasoning in Private Law' (2015) 35 OJLS (forthcoming)

Coote B, 'Contract Damages, Ruxley, and the Performance Interest' (1997) 56 CLJ 537

Cunnington R, 'Should Punitive Damages be Part of the Judicial Arsenal in Contract Cases?' (2006) 26 LS 369

Dharmananda K and Firios L, 'Penalties Arising Without Breach: The Australian Apogee of Orthodoxy' [2013] LMCLQ 145

Douglas J, 'England as a Source of Australian Law: For How Long?' (2012) 86 ALJ 333

Edelman J, 'A "Fusion Fallacy" Fallacy?' (2003) 119 LQR 375

Farber DA, 'Contract Law and Modern Economic Theory' (1983) 78 Nw UL Rev 303

Friedmann D, 'The Efficient Breach Fallacy' (1989) 18 JLS 1

Furmston MP, 'Contract Planning: Liquidated Damages, Deposits and the Foreseeability Rule' (1991) 4 JCL 1

Gardner J, 'What is Tort Law For? Part 1: the Place of Corrective Justice' (2011) 30 Law & Phil 1

Goetz CJ and Scott RE, 'Liquidated Damages, Penalties and the Just Compensation Principle: Some Notes on an Enforcement Model and a Theory of Efficient Breach' (1977) 77 Colum L Rev 554

Goff R, 'Rewriting Contracts and the Court of Appeal' (1961) 24 MLR 637

Ham A, 'The Rule Against Penalties in Contract: An Economic Perspective' (1990) 17 MULR 649

Harder S, 'The Relevance of Breach to the Applicability of the Rule Against Penalties' (2013) 30 JCL 52

Harris D, 'Incentives to Perform, or Break Contracts' (1992) 45 CLP 29

Heydon JD, 'Limits to the Powers of Ultimate Appellate Courts' (2006) 122 LQR 399

Hoffmann L, '*The Achilleas*: Custom and Practice or Foreseeability?' (2010) 14 Edin LR 47

Hohfeld WN, 'Some Fundamental Legal Conceptions as Applied in Judicial Reasoning' (1913) 23 Yale LJ 16

Holmes OW, 'The Path of the Law' (1897) 10 Har L Rev 457

Hudson AH, 'Penalties Limiting Damages' (1985) 101 LQR 480

Lanyon E, 'Equity and the Doctrine of Penalties' (1996) 9 JCL 234

McHugh M, 'The Growth of Legislation and Litigation' (1995) 69 ALJ 37

McHugh M, 'The Judicial Method' (1999) 73 ALJ 37

O'Farrell F, 'Challenging and Defending Liquidated Damages' (2004) 56 Amicus Curiae 20

O'Sullivan J, 'Lost On Penalties' (2014) 73 CLJ 480

Peden E, 'Forfeiture of Deposits: Where Law and Equity Collide?' (2012) 6 J Eq 161

Peel E, 'The Rule Against Penalties' (2013) 129 LQR 152

— — 'Unjustified Penalties or an Unjustified The Rule Against Penalties?' (2014) 130 LQR 365

Penner JE, 'The Analysis of Rights' (1997) 10 Ratio Juris 300

Posner RA, 'Some Uses and Abuses of Economics in Law' (1979) 46 U Chi L Rev 281

— — 'Legal Reasoning From the Top Down and From the Bottom Up' (1992) 59 U Chi L Rev 433

Rickett CEF, 'Punitive Damages: the Pulse of Equity' (2003) 77 ALJ 496

Rowan S, 'For the Recognition of Remedial Terms Agreed Inter Partes' (2010) 126 LQR 448

— — 'Reflections on the Introduction of Punitive Damages for Breach of Contract' (2010) 30 OJLS 495

Schwartz A, 'The Myth that Promisees Prefer Supracompensatory Remedies' (1990) 100 Yale LJ 369

Simpson AWB, 'The Penal Bond with Conditional Defeasance' (1966) 82 LQR 392

Smith L, 'Relief Against Forfeiture: a Restatement' (2001) 60 CLJ 178

— — 'Restitution: The Heart of Corrective Justice' (2001) 79 Tex L Rev 2115

Smith SA, 'Review: Taking Law Seriously' (2000) 50 UTLJ 241

Stevens R and McFarlane B, 'In Defence of *Sumpter v Hedges*' (2002) 118 LQR 569

Varghese RT, 'Penalties and Agreed Damages' (1992) 8 BCLRS 270

Wedderburn KW, 'Hire-Purchase—Penalties—Freedom of Contract' [1961] CLJ 156

Weinrib EJ, 'Corrective Justice in a Nutshell' (2002) 52 UTLJ 349

— — 'Deterrence and Corrective Justice' (2002) 50 UCLA L Rev 621

Zakrzewski R, 'The Nature of a Claim on an Indemnity' (2006) 22 JCL 54

3 Law Commission reports and papers

Law Commission, *Penalty Clauses and Forfeiture of Monies Paid* (Law Com WP No 61, 1975)

— — *Aggravated, Exemplary and Restitutionary Damages* (Law Com No 247, 1997)

Scottish Law Commission, *Report on Penalty Clauses* (Scot Law Com No 171, 1999)

Victorian Law Reform Commission, *Liquidated Damages and Penalties* (VLRC DP No 10, 1988)

4 Online resources

Oxford Dictionaries <http://www.oxforddictionaries.com/>