



University
of Glasgow

MacNeil, I. (2009) *Uncertainty in commercial law*. Edinburgh Law Review, 13 (1). pp. 68-99. ISSN 1364-9809

<http://eprints.gla.ac.uk/37965/>

Deposited on: 02 April 2012

Uncertainty in Commercial Law

Iain MacNeil^{*}

A. INTRODUCTION

B. THE MEANING OF LEGAL CERTAINTY

- (1) Predictability and consistency
- (2) Rules, principles and standards
- (3) Certainty and rule types
- (4) Certainty and vagueness of definition
- (5) Certainty and efficiency

C. PROXIES FOR LEGAL UNCERTAINTY

D. RESPONSES TO UNCERTAINTY

- (1) Eliminating or reducing uncertainty through contract terms
- (2) Allocating or transferring risk arising from legal uncertainty through contract terms
- (3) Creative compliance
- (4) Clearance, guidance and legal opinions

E. THE PERSISTENCE OF UNCERTAINTY

F. THREE EXAMPLES OF PERSISTENT UNCERTAINTY

- (1) The nineteenth century law on corporate personality and limited liability
- (2) The duty of disclosure in insurance
- (3) Material breach in contract law

G. CONCLUSIONS

A. INTRODUCTION

Considerable attention has recently been devoted to examining the relationship between law and uncertain future events. Particularly prominent in that debate has been the role of law in managing and allocating risk. In the field of contract

^{*} Alexander Stone Professor of Commercial Law, University of Glasgow. A draft of this paper was presented at a symposium on legal certainty held at Glasgow University in September 2006, funded by the Clark Foundation for Legal Education. I am grateful to the participants at that symposium, to Professor Doreen McBarnet and to the anonymous referee for comments.

and commercial law there now exists an extensive literature examining the incomplete nature of contracts and the extent to which contracting parties rely on mechanisms other than law in dealing with contingency. Within that literature, the focus is primarily on uncertainty arising from the unpredictability of events or developments that lie in the future. Much less attention has focused on the problem of legal uncertainty.

This article attempts to develop a working definition of legal uncertainty and to examine its effect in commercial law. It is based on the premise that lawyers in the commercial world have to deal with and contract around not only future contingencies but also legal uncertainty. The article attempts to explain why uncertainty may persist over relatively long periods of time in commercial law, despite the perceived need for legal certainty to facilitate commercial transactions. It also attempts to explain, both in general terms and by reference to specific examples, how uncertainty can be limited or removed. Underlying much of the discussion is the argument that there is a close link between the effectiveness of responses to uncertainty and the persistence of uncertainty over time.

B. THE MEANING OF LEGAL CERTAINTY

(1) Predictability and consistency

It is often said that business activity is facilitated by legal certainty.¹ This implies that the law should be predictable and should treat similar cases consistently.² Predictability allows those subject to the law to judge the law's reaction to their conduct (such as when making a choice between legal and illegal action).³ The content of any particular law can only be understood in a meaningful way if its application to particular circumstances can be predicted.⁴ Consistency is closely related to predictability, but focuses less on the outcome of a particular adjudication by comparison with what the law is generally understood to be and more on the relative outcomes of different adjudications that apply the same law.

1 See e.g. L S Sealy and R J A Hooley, *Commercial Law: Text, Cases and Materials* (2003) 10: "Businessmen have special needs . . . they require the decisions of the courts on commercial issues to be predictable so that they know where they stand."

2 In this sense, legal certainty is not directly concerned with the substance of the law, a point made by Lord Mansfield in *Vallejo v Wheeler* (1774) 1 Cowp 143 at 153: "In all mercantile transactions the great object should be certainty: and therefore, it is of more consequence that a rule be certain, than whether the rule is established one way or the other. Because speculators in trade then know what ground to go upon."

3 See generally R Goode, "The philosophy and concepts of commercial law" (1988) 14 *Monash Law Review* 135.

4 In some instances, predictability has been characterised as serving legal advisers as much as persons subject to the law. See e.g. *O'Neill v Phillips* [1999] 2 All ER 961 at 967 per Lord Hoffmann.

Legal uncertainty differs from legal risk in that the latter focuses on the chances of being sued or being the subject of a claim or the possibility that a technical defect in a transaction will result in loss.⁵ In that sense risk is a much broader concept than uncertainty because it arises even where the law is clear. Legal uncertainty is that subset of risk which focuses on circumstances in which it is the lack of clarity in the law that poses risks for markets, transactions and legal structures. Two forms of legal uncertainty are important for commercial law. The first is where some doubt remains unresolved over a substantial period of time. This may arise because the doubt is not the subject of litigation before a court.⁶ It may also arise from the process of clarification, testing and interpretation that is represented by case law that does not claim to change the existing law but which nevertheless carries implications for its reliability.⁷ While this form of uncertainty may not appear to pose a substantial risk for the commercial world, it does raise issues regarding the extent to which precedent can provide a reliable guide for future transactions.⁸ For example, the seminal company law case *Salomon v Salomon*⁹ did not overturn any of the previous authorities yet is widely regarded as marking an important turning point in the recognition of limited liability and separate corporate personality.¹⁰ From the perspective of the law in practice, uncertainty may be reduced or removed by expert consensus. This is likely to be important in areas where a relatively small community of legal advisers is engaged in innovative transactions which have not been tested in the courts but rely for legal certainty on consensus among those advisers, perhaps combined with the development of standard documentation.¹¹

The second form of uncertainty occurs when a “settled view” of the law is subsequently changed by the courts.¹² Although this happens only infrequently,¹³

5 See e.g. R McCormick, *Legal Risk in the Financial Markets* (2006) 10 for a definition of legal risk.

6 One explanation of how this may occur is offered in B.(5) below.

7 See S Waddams, *Dimensions of Private Law* (2003) 206 regarding the distinction between clarifying or applying an existing rule and creating a new rule.

8 For discussion of how this may operate when courts choose to distinguish or “not follow” previous authority, see R H S Tur, “Time and law” (2002) 22 OJLS 463.

9 [1897] AC 22, discussed at F. (1) below.

10 Similarly, the decision of the Court of Appeal in *Great Peace Shipping Ltd v Tsavliris Salvage (International) Ltd* [2002] QB 679 dealt with the problem of reconciling the respective approaches of common law and equity to mistakes in law by “disapproving” the leading decision made under the equitable jurisdiction, *Solle v Butcher* [1950] 1 KB 671.

11 The growth of securitisation in recent years is an example of such a process.

12 Subsequent change effected by statute does not pose a risk as legislation does not have retrospective effect.

13 A modern example is *In re Spectrum Plus Ltd (In Liquidation)* [2005] 2 AC 680. In that case the House of Lords provided a justification for overturning settled law on the basis that (a) those who rely on a decision at first instance must be taken to be aware of the possibility of it being overturned and (b) the statutory priority given to the holder of a fixed charge over preferential creditors must be regarded as superior to the need to preserve legal certainty through upholding prior case law.

it can have a serious impact. For example, *Morgan Guaranty*¹⁴ and *Kleinwort Benson*¹⁵ established that it is possible to recover a payment made under a mistake of law.¹⁶ The rationale is that, while there may have been no mistake of law at the time the obligation to make the payment was concluded, re-interpretation of the law in a later case applies retrospectively, opening up the possibility of a case based on mistake.¹⁷

The significance of legal certainty, and the cause of much of the debate as to the extent to which it is achieved,¹⁸ lies in its link with legitimacy. Kress explains that link in the following terms:¹⁹

Indeterminacy matters because legitimacy matters. Many legal scholars hold that the legitimacy of judicial decision-making depends upon judges applying the law and not creating their own. They claim that judicial decisions are legitimate only if judges are constrained either completely or within narrow bounds.

On this view, legal certainty has a role to play in ensuring that courts resolve disputes according to established legal norms and, in that sense, act in a legitimate manner. While not the source of legitimacy, legal certainty is nevertheless closely linked with it, through its role in limiting arbitrary judicial decision-making. A related issue is the extent to which the subjects of law are able to gain access to and understand the process of legal reasoning through which cases are resolved: that may exert just as powerful an influence on the exercise of rights and the resolution of disputes as the content of the law itself.²⁰

Legal certainty can also be linked with informed decision-making and efficient markets. It is now generally accepted that one of the conditions required for

14 *Morgan Guaranty Trust Co of New York v Lothian Regional Council* 1995 SC 151.

15 *Kleinwort Benson Ltd v Lincoln City Council* [1999] 2 AC 349. For criticism, see M Bridge, "Restitution and retrospective law: *Kleinwort Benson v Lincoln City Council*" (1999) 14 *Butterworths Journal of International Banking and Financial Law* 5.

16 In *Kleinwort Benson* Lord Hope (at 410B) distinguished a mistake of law from the presence of some doubt over the law. Moreover, as is made clear by *Brennan v Bolt Burdon* [2005] QB 303, it is a matter of interpretation whether a party to an agreement or compromise has surrendered rights of which that party is unaware or could not be aware because they have not yet been recognised by the courts. See also *Mahmud v BCCI* [1998] AC 20.

17 *Kleinwort Benson* at 378 per Lord Goff. It remains something of a mystery why so few cases based on mistake of law come before the courts. In some areas, such as insurance law, where the law has been in flux in recent years, there would seem to be scope for claims paid or refused on the basis of the law at the time of payment to be re-opened on the basis of a mistake of law. While *Brennan v Bolt Burdon* [2005] QB 303 can be viewed as restricting the scope for holding a contract void as a result of mistake, it does make clear that a compromise of a claim can also be held void as a result of a mistake of law. A rare example of the application of the principle is *BCCI v Ali* [2002] 1 AC 251.

18 See e.g. K Kress, "Legal indeterminacy" (1989) 77 *Cal L Rev* 283.

19 Kress (n 18) at 285.

20 See e.g. A Halpin, *Reasoning with Law* (2001) 104, arguing that there may be questions posed regarding the legitimacy of an unelected judiciary using the "art of legal reasoning" to determine the legal fate of citizens.

markets to operate efficiently is that the participants in those markets should be able to make fully-informed decisions.²¹ In some cases – as for example with disclosure obligations in insurance or for share prospectuses – the law makes available information which is important for decisions. In other cases, the making of a decision encompasses a view as to the law regarding a particular issue, so that legal uncertainty will limit the extent to which the decision can be fully-informed. For example, a decision to rescind a contract presupposes knowledge of the law of contract. A party exercising the right of rescission bears the risk that the act will be considered to be (unjustified) repudiation rather than (justifiable) rescission.²² Thus, uncertainty in the law is likely to increase risk and to operate as a disincentive to engage in market transactions.

A further aspect of uncertainty is cost. This includes higher litigation and judicial system costs, increased unlawful activity and decreased lawful activity, and replacement costs.²³ Uncertainty is likely to affect the incidence of unlawful activity because it blurs the boundary between what is permitted and what is not. These costs have a temporal dimension in that it is likely that uncertainty will diminish over time as courts interpret the law and reforms are enacted, but that process itself carries with it substantial costs, particularly if the resolution of uncertainty is not channelled in a manner designed to produce the most informed result possible.²⁴

(2) Rules, principles and standards

The extent to which the law is certain is related to the manner in which it is structured and distributed as between principles and rules. Principles are broad statements that apply to a range of acts and, in Common Law systems, tend to develop through a series of cases rather than being established in a single case.²⁵ Rules are more specific and relate to particular acts. As Braithwaite observes, it is often assumed in legal scholarship that tightly specified legal rules increase legal certainty,²⁶ an approach which favours rules over principles. The argument

21 See e.g. A Ogas, *Regulation, Legal Form and Economic Theory* (1994) 38-41.

22 This particular issue is examined in more detail in part F.3 below.

23 See generally G Maggs, "Reducing the costs of statutory ambiguity: alternative approaches and the federal courts study committee" (1992) 29 *Harv J on Legis* 123.

24 See e.g. L Kaplow, "Rules versus standards: an economic analysis" (1992) 42 *Duke L J* 557, referring to the possibility of many adjudications having been made before an authoritative precedent emerges.

25 See generally J Raz, "Legal principles and the limits of law" (1971-72) 81 *Yale L J* 823 at 839.

26 J Braithwaite, "Rules and principles: a theory of legal certainty" (2002) 27 *Australian Journal of Legal Philosophy* 47 at 51, citing C S Diver, "Regulatory precision", in K Hawkins and J Thomas (eds), *Making Regulatory Policy* (1990).

is made in the following terms by Raz:²⁷

Since the law should strive to balance certainty and reliability against flexibility, it is on the whole wise legal policy to use rules as much as possible for regulating human behaviour because they are more certain than principles and lend themselves more easily to uniform and predictable applications.

As Braithwaite argues, however, favouring rules in pursuit of legal certainty may be a strategy that is doomed to failure. Admittedly, when the type of action to be regulated is simple and stable (not changing unpredictably across time) and does not involve huge economic interests, rules regulate with greater certainty than principles.²⁸ But with complex actions in changing environments where large economic interests are at stake, principles, Braithwaite argues, are more likely to enable legal certainty than rules. Uncertainty will result from a process by which the legal advisers of large economic interests (“wealthy legal game players” in Braithwaite’s terminology) engage in a process which constantly challenges and attempts to redefine “grey areas”²⁹ of the law. This type of activity has been variously termed “creative compliance”³⁰ and “compliant non-compliance”³¹ and is particularly evident in tax law, company law and financial regulation. In essence, both involve the pursuit of legal strategies that pay lip service to the detail of regulatory rules but nevertheless pursue objectives that run contrary to their spirit. It is the fragmentation of the spirit of the rules through the pursuit of certainty, itself often the result of the “game” by which wealthy legal game players test the rules, that makes such a strategy possible.

A similar approach is adopted by McBarnet and Whelan in their analysis of the regulation of accounting standards. They argue that an approach based on principles, while offering an ostensibly attractive method of policing “creative accounting”, is open to challenge through several techniques, which they collectively refer to as “looking for loopholes”. The central point of their argument is that principles do not necessarily prohibit creative accounting more successfully than rules because there are compliance strategies (“creative compliance”) which can respond just as effectively to principles as to rules. Thus, normative claims about the superiority of principles over rules or *vice versa* are inherently suspect and may themselves represent strategies to gain political traction in determining

27 Raz (n 25) at 841.

28 Braithwaite (n 26) at 54.

29 Or in Hart’s terminology the “penumbra” of legal rules: see H L A Hart, *The Concept of Law*, 2nd edn (1994) 251.

30 D McBarnet and C Whelan, “The elusive spirit of the law: formalism and the struggle for legal control” (1991) 54 MLR 848.

31 See D McBarnet and C Whelan, *Creative Accounting and the Cross-eyed Javelin Thrower* (1999).

the direction of regulatory development.³² I return to this issue below in the context of responses to uncertainty.³³

Legal uncertainty can also be approached from the perspective of the balance in law between rules whose content is defined *ex ante* and standards whose content is defined *ex post* (usually by a court or other adjudicative body). For example, in the field of corporate law, a choice can be made in controlling self-dealing on the part of directors between setting rules that determine how and when a director is permitted to enter into contracts in which they have a conflict of interest or allowing the matter to be settled *ex post* by applying the overriding principle of fiduciary duty.³⁴ As was noted above in respect of principles versus rules, there is a similar tendency to assume that rules generate greater legal certainty than standards, but that may not necessarily be the case. Kaplow, for example, argues that there are circumstances in which standards offer a more certain approach than rules and *vice versa*, and that there is therefore no reason to assume that one technique is superior to the other.³⁵ Of course, to the extent that the application of standards creates binding precedent, the distinction between standards and rules is greatly reduced, but the process by which binding precedent is created may take time, create uncertainty (e.g. because of the limited applicability of the precedent) and prove costly.³⁶

(3) Certainty and rule types

Braithwaite's argument against over-reliance on rules as opposed to principles is formulated primarily by reference to regulatory rules.³⁷ But while it is true that the scope of regulation has increased greatly in recent years, it remains the case that much of the law relating to commercial activity does not take the form of regulatory (mandatory) rules. From the perspective of transactions, a great deal of contract law (the core of transactional law) comprises default rather than mandatory rules. The same is true of the law relating to business organisations, both in respect of partnerships and companies. Therefore, a broader view of the consequences of uncertainty should take account of the wider range of rules that are relevant for commercial activity.

32 See in this regard, L A Cunningham, "A prescription to retire the rhetoric of 'principles-based systems' in corporate law, securities regulation and accounting", available at <http://ssrn.com/abstract=970646>; D Kershaw, "Evading Enron: taking principles too seriously in accounting regulation" (2005) 68 MLR 594.

33 See D.(3) below.

34 See e.g. H Hansmann and R Kraakman, *The Anatomy of Corporate Law* (2004).

35 Kaplow (n 24) at 585-590.

36 Kaplow (n 24) at 614-615.

37 Braithwaite (n 26) at 54.

Default rules serve a range of purposes. First, they represent a (limited) response to the problem of incomplete contracting by filling in gaps in the contract.³⁸ For example, the law relating to implied contract terms can be described as a default rule which will, in some circumstances, recognise as part of a contract a term that was not expressly agreed between the parties.³⁹ Second, from a law and economics perspective, default rules are envisaged as promoting efficiency by allowing the parties to economise on transaction costs.⁴⁰ This is linked with the “gap filling” function in that it is the role of default rules in filling gaps that allows the contracting parties to limit their agreement to its essential features.

The presence of uncertainty in default rules does not lead to the same outcome as in the case of mandatory rules because there is no obligation to obey default rules. Indeed, it is envisaged that the contracting parties and society generally may be better served by amendment or disapplication of default rules rather than by their acceptance.⁴¹ If a default rule is uncertain, it cannot properly fulfil its function as a gap filler. This is because uncertain default rules create risks for the contracting parties when they decide what must be expressly agreed and what can safely be left to the general law, with the likely result that increasing use will need to be made of express terms. Suppose for example that the default rule in contract law is that a delay in performance does not constitute grounds for rescission. If it is not clear what constitutes a delay by comparison with refusal to perform, it may be that one side will want to include an express right to rescind following a delay of ten days so as to avoid the risk of open-ended delay with no remedy. In other words, the parties’ attitude to the express terms of the contract will be influenced by the formulation (including legal certainty) of the relevant default rules.⁴²

“Enabling” rules are also important in commercial law, particularly in the field of business organisations. These are rules which enable the doing of what would not otherwise be possible. For example, the rules of company law that allow a registered company to be created make possible an outcome (the creation of an entity distinct from the shareholders) that could not be achieved (despite much

38 See generally I Ayres and R Gertner, “Filling gaps in incomplete contracts: an economic theory of default rules” (1989) 99 *Yale L J* 87.

39 Its operation can be excluded by an “entire agreement clause”, which has the effect of limiting the contract to terms expressly agreed. See generally I MacNeil, “Excluding liability for misrepresentation” (1999) 3 *SLPQ* 226.

40 For an overview of the significance of transaction costs, see O Williamson, *The Economic Institutions of Capitalism* (1985) 15-43.

41 See Ayres & Gertner (n 38); C Goetz and R Scott, “The limits of expanded choice: an analysis of the interactions between express and implied contract terms” (1985) 73 *Cal L Rev* 261.

42 See, for a comparative analysis of this issue and recognition of the “uncertainty inherent in the English approach”, J E Stannard, “In the contractual last chance saloon: notices making time of the essence” (2004) 120 *LQR* 137.

ingenious effort)⁴³ by other legal devices. Uncertainty in enabling rules limits the extent to which the relevant activity or legal structure will be enabled, because the requirements that must be met to gain the benefit of the rule will not be clear. In the example just given, uncertainty is likely to limit the attraction of the company form, because there would remain doubt over the scope of the separate legal personality rule, and shareholders would fear that the company's creditors might look to them for satisfaction of the company's debts.

(4) Certainty and vagueness of definition

Uncertainty may arise when the meaning of a term around which a rule is formulated is not settled by accepted usage. An example given by Halpin is the use of the word "bald" in a legal rule.⁴⁴ Three forms of vagueness relating to this word can be distinguished:

- (a) Linguistic infelicity. This follows from the observation that it is possible to maintain that a person is neither bald nor "not bald".
- (b) Conflicting usage. This may result when two parties (e.g. contracting parties) use the term in different ways, irrespective of how the rest of the world uses the term.
- (c) Unsettled usage. This occurs when there is no relevant agreement in the relevant community on the meaning of the word.

In addressing the issue of whether it is possible for such a term to remain vague, Halpin concludes that it can remain vague "for so long as there may arise novel situations for which the application of the term can be considered or situations over which conflicting usage has not conformed to an accepted usage."⁴⁵ The removal of uncertainty is viewed as a process by which an inherent tendency towards linguistic vagueness is resolved as the courts establish settled usage.

Linguistic uncertainty is a sub-set of the first type of uncertainty referred to earlier, which encompasses the three forms of vagueness mentioned above. A characteristic of this type of uncertainty is that the process of establishing settled usage, while not representing legal change, may well have substantial effects for those who subscribed to the discredited version of conflicting usage.⁴⁶

43 For a historical perspective see P Ireland, "The rise of the limited liability company" (1984) 12 *International Journal of the Sociology of Law* 239.

44 Halpin, *Reasoning with Law* (n 20) 97.

45 Halpin, *Reasoning with Law* (n 20) 99.

46 See e.g. *Sharp v Thomson* 1997 SC (HL) 66 regarding the meaning of "property" for the purposes of s 464 of the Companies Act 1985 and the effect of this decision on those who subscribed to the (discredited) view that it comprised all property in respect of which title was recorded in the Land Register or Register of Sasines.

In this sense, the legal process of establishing settled usage poses risks for existing contracts and other legal obligations, even if it carries with it benefits in terms of legal certainty for the future.

(5) Certainty and efficiency

Regardless of whether particular legal rules are efficient, it can be argued that the common law has an inherent tendency to evolve in the direction of efficient rather than inefficient rules. Put another way, the argument is that the common law will favour rules that impose fewer costs, and in that sense will tend to promote aggregate social welfare. A variant of that argument is Posner's view that the common law provides incentives to judges and other procedural devices which ensure efficient outcomes in individual cases and therefore also in the aggregate.⁴⁷ An alternative view, which has attracted considerable support, explains the evolution towards efficient rules by reference to the incentives of parties to litigate or settle outside court rather than by reference to the motivations or tendencies of judges. According to this view, the evolution of the common law depends on cases which are litigated and parties will only litigate when a rule is inefficient and they have significant economic interests at stake.⁴⁸ Over time, the law will come to comprise a group of settled and efficient rules which are subject to little or no litigation as well as a group of inefficient rules which continue to be litigated. This evolutionary process can be viewed as independent of any tendency on the part of the judiciary to favour efficient rules, as re-litigating inefficient rules will create ongoing pressure for them to be changed. Rubin comments that "Intelligent judges may speed up the process of attaining efficiency; they do not drive the process."⁴⁹

It has been argued that there is no direct link between certainty and efficiency because there is no reason to believe that, in general, efficient rules are more certain than those which are not efficient.⁵⁰ The characteristic which makes a rule efficient is the cost which it imposes on parties and the effect which that cost has on the allocation of resources: in that sense a rule subject to some degree of uncertainty can still be more efficient than one to which no uncertainty seems to attach. However, taken to its extreme, that view cannot be correct as there must come a point at which a rule becomes so uncertain that it cannot allocate costs because it is not clear what the rule requires.⁵¹

47 R A Posner, *Economic Analysis of Law*, 7th edn (2007) 249-253, 555-639.

48 See e.g. P H Rubin, "Why is the common law efficient?" (1977) 6 J Leg Stud 51; G L Priest, "The common law process and the selection of efficient rules" (1977) 6 J Leg Stud 65.

49 Rubin (n 48) at 55.

50 As argued by Priest (n 48) at 68.

51 E.g. an uncertain rule which purports to determine when auditors are liable for inaccurate information in audited accounts cannot effectively allocate the burden of prevention of loss to auditors or third parties who rely on the information.

A better view is that certainty does carry implications for the efficiency explanation of the process of (common law) evolution in that it seems likely to affect the propensity of parties to a dispute to litigate, and thereby assist the development of the common law, rather than to settle. If a rule appears certain to both sides there will be a reduced tendency to litigate (even if the rule is efficient) because the probability of success is low. If, conversely, uncertainty is present, the tendency to litigate will depend on the likelihood of a successful outcome. In that sense, certainty affects the chances of litigation, for rules embedded in a legal system with certainty are unlikely to be litigated.⁵² Naturally, this is not the only factor on which a decision to litigate is based: other factors include the interests of the parties in the precedent value of the case for their business⁵³ and the relative resources they are able to invest in litigation.

C. PROXIES FOR LEGAL UNCERTAINTY

So far the discussion has avoided the issue of how to identify uncertainty in the law. That issue poses a problem because it is widely accepted that there will always be some degree of uncertainty—in Hart’s terminology any legal provision will be associated with a core of certainty and a penumbra of uncertainty.⁵⁴ Thus, it is inevitable that there will be a process of testing, defining and redefining legal provisions so as to expand the core and limit the penumbra, and much litigation can be viewed in this light.⁵⁵ While this is often no more than the normal process of legal development, it is possible for uncertainty to be present in a more pervasive manner, such as when the degree of uncertainty indicates that the core itself may be indeterminate. Kress approaches this issue from the perspective that “Law is indeterminate where the correct theory of legal reasoning fails to yield a right answer or permits multiple answers to legal questions”.⁵⁶ On this approach,

52 In those cases, there would not even seem to be scope for a dispute between informed parties.

53 Rubin (n 48) at 53 cites the example of the interest of insurers in precedents establishing liability for accidents, which have substantial implications for their business.

54 Cf Hart, *Concept of Law* (n 29). See also T Endicott, *Vagueness in Law* (2000), arguing that vagueness and indeterminacy are essential features of the law. Critical legal theorists have expressed the inevitable indeterminacy of law in its most extreme form. See e.g. M Kelman, *A Guide to Critical Legal Studies* (1987) 258, arguing that “All rules will contain within them, deeply embedded, structural premises that clearly enable decision makers to resolve particular controversies in opposite ways.”

55 But not all. See e.g. Waddams, *Dimensions of Private Law* (n 7) regarding the role of litigation in setting the conceptual framework for the solution of legal problems rather than just simply applying rules.

56 Kress (n18) at 283, 325. Kress argues against the view of critical legal scholars, that indeterminacy is widespread and that law therefore lacks legitimacy, on three grounds: (a) critical legal scholars unnecessarily limit the premises that can be employed in legal argument by focusing on black-letter rules and excluding standards such as principles, policies, legal culture and social context; (b) they focus excessively on deductive logic and deny legitimacy to other techniques such as analogy; and (c) they assume that the existence of conflicting principles leads to indeterminacy in the absence of explicit metaprinciples that resolve those conflicts.

what is at issue is whether the law can supply a single correct answer within the “core”, not the categorisation or characterisation of the rules and reasoning by which the answer is reached.

Identifying instances in which the law is uncertain in this sense requires some form of proxy⁵⁷ to be used for uncertainty. One proxy that has been used, in the context of statutory provisions, is that a legal provision is uncertain when it would be litigated by a lawyer.⁵⁸ While this definition might seem overly broad, it has the merit of linking uncertainty with the cost of its resolution, since a rationalisation of the litigation process would lead to the conclusion that the expected benefit of resolution for the litigant would be greater than either the cost of the litigation or any other solutions that might be available to resolve uncertainty. In other words, the definition focuses on the practical significance of uncertainty, and assumes that it is only uncertainty which has a substantial cost which will be tested by litigation.

Another test for uncertainty is the extent to which dissenting judgments are made in appellate courts.⁵⁹ This has the merit of focusing attention on disagreement over “settled law”. It is based on the premise that disagreements in the appellate courts are likely to be in respect of issues that are particularly significant from a certainty perspective as they provide a strong indication that there is no “settled law” on the matter. Kress finds relatively few instances of such judgements and concludes that this is an indicator that the law is, in the main, predictable in its application.⁶⁰

Both of these proxies for legal uncertainty suffer from the deficiency that they deal with cases that have been litigated, and so ignore the many instances in which uncertainty is resolved outside the litigation process.⁶¹ For the purposes of this article, therefore, uncertainty is taken to exist in circumstances in which one or more of a range of actions (which are in essence responses to the presence of uncertainty) is routinely employed. Admittedly this results in a less clearly defined set of situations, because there may be scope for disagreement over whether an uncertainty-resolving strategy has in fact been deployed or whether it

57 A proxy in this sense means a working definition that is a close substitute for the underlying concept. It is an approximation for the purposes of exposition.

58 Maggs (n 23).

59 See Kress (n 18) at 324-325 for statistics. Kress uses this proxy to support his more general proposition that “Law is indeterminate when the correct theory of legal reasoning fails to yield a right answer or permits multiple answers to legal questions” (320).

60 Kress (n 18) at 325, noting that dissenting judgments are nevertheless not a precise measurement of judicial disagreement because dissent is not always noted.

61 See J M Lipshaw, “Contingency and contracts: a philosophy of complex business transactions” (2005) 54 *DePaul L Rev* 1073; S Macaulay, “Non-contractual relations in business: a preliminary study” (1963) 28 *American Sociological Review* 55.

is a significant part of the legal arrangement of which it forms part. However, the benefit of this approach is that it enables uncertainty to encompass the many cases in which litigation is avoided because an appropriate response is otherwise available. Moreover, as will be argued later, this approach allows the development of a view as to the overall significance of uncertainty for commercial transactions.

This proxy for uncertainty encompasses both the first and the second forms of uncertainty referred to earlier.⁶² However, it does so in a manner which might best be described as *ex ante*: in other words, it refers to the possibility that the parties will judge the law to be sufficiently uncertain that a court adjudicating on the matter would conclude that one of these forms of uncertainty was present. This differs from the proxy of dissenting judgements in appellate courts which, by identifying uncertainty *ex post* by reference to the adjudication of the court, ignores the manner in which the parties negotiated and agreed the transaction in question. The proposed test thus recognises that uncertainty is often identified at the planning stage in transactions, long before a dispute has actually arisen.

Another feature of the proposed test is that it can help to explain the persistence of uncertainty over time. If, as mentioned earlier, there is a general demand for commercial law to be certain, it can be expected that one of the main functions of litigation is to resolve uncertainty. The process of “testing the limits” and looking for “loopholes”, described by McBarnet and Braithwaite in respect of regulatory rules,⁶³ is likely to be replicated in the sphere of private law by litigation as contracting parties aim to extract the maximum advantage from particular interpretations of the rules.⁶⁴ To the extent that litigation is avoided by the use of alternative mechanisms, it may be possible to argue that uncertainty in the law persists over a longer period of time because the issue is not litigated. This is an argument to which I will return after considering the potential responses to uncertainty.⁶⁵

D. RESPONSES TO UNCERTAINTY

There appear to be at least four distinct responses to the presence of legal uncertainty in commercial transactions.

⁶² See B.(1) above.

⁶³ See McBarnet & Whelan (n 30) pt III.

⁶⁴ E.g. see the discussion in *In re Spectrum Plus Ltd (in liquidation)* [2005] 2 AC 680, where the impact of differing interpretations of the essential characteristics of a fixed charge carried significant implications for different groups of creditors with competing interests.

⁶⁵ See E below.

(1) Eliminating or reducing uncertainty through contract terms

One response is to include specific provisions in a contract. A common example is the identification of circumstances which will give rise to the right to rescind. This process also occurs more generally when contracting parties tailor default rules to suit their own circumstances: prime examples can be seen in the formation of contracts of sale and the constitutions of companies, where default rules are frequently amended by the parties.⁶⁶ While the primary aim is to adapt rules to the particular circumstances of the parties, this also serves the purpose of eliminating legal uncertainty. The process of adaptation, however, exposes parties to the risk that their tailored version will be interpreted by the courts differently to what they (either in common or individually) expected.⁶⁷

(2) Allocating or transferring risk arising from legal uncertainty through contract terms

A transfer of risk arising from legal uncertainty could occur in two circumstances: first, where both sides are aware of the risk and the terms of the transaction make it attractive for one side to bear it;⁶⁸ second, where the presence of the risk is known only to one side, which is able to use its superior knowledge to transfer the risk to the other side.⁶⁹ While this process bears some similarity to that described in the previous section, the focus is on one side bearing the risk of legal uncertainty rather than both sides working together to remove uncertainty for their mutual benefit.⁷⁰ It also differs from the process of allocating risk in respect of uncertain future events. As well as being transferred between

66 The process of “tailoring” may also occur in an institutional setting: see e.g. V Gessner and A Cem Budak, *Emerging Legal Certainty: Empirical Studies on the Globalization of Law* (1998) for a discussion of how non-state institutions help to resolve legal uncertainty in transnational business transactions. While this process may focus on developing terms that are described as “standard” for transnational transactions, it inevitably represents a form of “tailoring” of some national rules that might otherwise apply.

67 I return to this issue in more detail at F.(3) below.

68 This could arise for example from the pricing of the contract.

69 For example, an astute seller of land to a developer is aware that there is some legal uncertainty in respect of a rule of planning law which may prevent the grant of planning permission over the land: a developer who is not aware of the uncertainty may be willing to purchase the land at a value which attaches too high a probability to planning consent being given.

70 This strategy may also be employed to allocate regulatory risk. See D McBarnet, “Transnational transactions: legal work, cross-border commerce and global regulation”, in M Likosky (ed), *Transnational Legal Processes* (2002), discussing the use of contractual terms in public offers of shares in Russia in respect of the liability of the issuer for the accuracy of its share register. Such terms were made necessary by the incomplete system of public regulation of share registers, which created risks in the recognition and transfer of legal title to shares.

contracting parties, such risks can also be transferred to a third party through insurance. However, the range of insurable risks does not include legal risk and so it follows that risks arising from legal uncertainty fall to the parties themselves. In particular, following the courts' recognition of the right to recover payments made under a mistake of law,⁷¹ there is a form of sleeping uncertainty inherent in every transaction in that the possibility of settled law being overturned is always present, even if it is not a regular occurrence.

(3) Creative compliance

Creative compliance involves compliance with the letter of the law while pursuing an objective which is contrary to its spirit.⁷² It also represents a response to legal uncertainty in that it attempts to exploit the presence of uncertainty for the benefit of the subject of regulation. In that sense it differs fundamentally from the two techniques referred to above because it is based on the premise that uncertainty may offer benefit. Three techniques for achieving creative compliance have been identified by McBarnet and Whelan:⁷³

- (a) Searching for loopholes. This involves the structuring of transactions or the performance of regulatory duties (such as financial reporting by companies) in a manner that aims to take advantage of gaps, exceptions or "overrides"⁷⁴ within the regulatory rules.
- (b) Falling outside the scope of regulation. This technique involves the structuring of transactions so as to fall outside the scope of regulated activity. A classic example was the development during the 1980s of off-balance sheet financing through the creation of "orphan subsidiaries" which did not fall within the (then) definition of a subsidiary in the Companies Act 1985. The result was that companies were able to devise structures that did not appear on their balance sheet, thereby frustrating the "true and fair" requirement imposed on directors in respect of a company's financial statements.⁷⁵

71 See B.(1) above.

72 See generally D McBarnet (ed), *Crime, Compliance and Control* (2004).

73 McBarnet & Whelan, *Creative Accounting* (n 31) chs 8-10.

74 McBarnet & Whelan, *Creative Accounting* (n 31) 11 cites the Companies Act 2006 s 393 in respect of company financial reporting. This allows the directors to override the specific requirements of the Companies Act if, in special circumstances, those requirements are inconsistent with the requirement to give a "true and fair view" of the financial position of the company.

75 McBarnet & Whelan, *Creative Accounting* (n 31) 121 notes that, ironically, there were no prosecutions undertaken by the DTI under s 234(5) of the Companies Act 1985 for omitting an "orphan subsidiary" from a group balance sheet. On the contrary, the fact that the only prosecution (the Argyll case, unreported) was for the wrongful *inclusion* of a subsidiary in group accounts only served to legitimise the practice of creating "orphan subsidiaries".

(c) Working to rule. In the words of McBarnet and Whelan:⁷⁶

This involves focusing literally and narrowly on the words of a rule, and working creatively on it . . . To pursue this strategy, however, one must find a rule to work to, the more precise and prescriptive the better.

One objective of working to rule may be to find a relevant detailed rule, compliance with which can be argued to represent compliance with a governing principle which has a higher normative value but which lacks direct applicability to the circumstances. In that sense, the certainty which detailed rules attempt to achieve is undermined by their very creation, which then sets in motion the process of working to rule. McBarnet and Whelan provide as an example the attempts made by companies to fall within the ambit of specific rules that implement the general principle that companies should report the substance and commercial effect of transactions.⁷⁷

Creative compliance may test the limits of the law, such as occurs regularly in relation to the distinction between (legal) tax evasion and (illegal) tax avoidance. It is likely to promote clarification of the law, as the process of successively finding and testing loopholes would suggest that the core of settled law is likely to expand over time.⁷⁸ But, as discussed below in the context of *Salomon v Salomon*,⁷⁹ it may also have the effect of delaying the emergence of settled law because the presence of uncertainty allows the subjects of regulation to achieve their objectives through creative compliance.

(4) Clearance, guidance and legal opinions

Some regulatory systems cope with uncertainty by the provision of “clearance” decisions, which confirm that a proposed transaction or conduct complies with regulatory rules. This model operates in EC competition law, allowing the Commission to find that articles 81 or 82 of the EU Treaty do not apply in particular circumstances.⁸⁰ A variant is where the system does not provide expressly for such decisions to be given but allows this on an informal basis. An example was the practice of issuing “comfort letters” that developed to enable the European Commission to deal with the many enquiries it received regarding

⁷⁶ McBarnet & Whelan, *Creative Accounting* (n 31) 137.

⁷⁷ McBarnet & Whelan, *Creative Accounting* (n 31) 146.

⁷⁸ However, to the extent that the law is in constant transition, the process of testing will not contribute substantially to an expansion of the settled core of law. Tax law in the UK may well be a case in point.

⁷⁹ [1897] AC 22. See F.(1) below.

⁸⁰ Commission Regulation 1/2003 OJ 2003 L1/1 art 10

compliance with article 81 of the EU Treaty (relating to cartels) without making a formal decision.⁸¹

Guidance provides a form of lower level of legal certainty than clearance. It can take the form of generic guidance that indicates how a regulator will interpret rules,⁸² or specific guidance which deals with individual circumstances. In the latter case, it may be difficult to distinguish guidance from clearance. While the emergence of guidance may be viewed as a positive development by those favouring detailed rules over general principles, it has also been characterised as a threat to anti-formalist approaches to regulation which focus on substance over legal form. This issue has been to the fore in the context of tax law, where guidance may be called for to support tax minimisation strategies, as well as in the field of accounting, where guidance may support creative compliance.⁸³

Legal opinions may also provide a form of legal certainty. They are relied on to a significant extent in financial markets and may even be integrated into regulatory requirements.⁸⁴ For example, the UK Financial Services Authority sets requirements for legal opinions given in connection with close-out netting agreements, which control netting of debts and set-off following the default of a market participant. These legal opinions are relied on not only by market participants in controlling their own business risks but also by the regulator in controlling the solvency of regulated firms.⁸⁵

E. THE PERSISTENCE OF UNCERTAINTY

Uncertainty in commercial law can persist over long periods of time, even in respect of quite basic rules.⁸⁶ The cause of such persistence may be the existence of the very mechanisms – the responses to uncertainty discussed in part D – by which uncertainty can be resolved, on the basis that the need to resolve uncertainty is then reduced or eliminated. If that is correct, it is possible to claim that the more successful the responses to uncertainty, the longer the uncertainty is likely to survive. In that sense, legal ingenuity, and perhaps reliance on non-legal measures, can be associated with an ability to tolerate uncertainty within the law governing commercial transactions and organisations. If an adequate response to uncertainty removes its damaging effect (at least for the transaction

81 This system operated under Regulation 17/62 1959-62 OJ Spec Ed 87 and came to an end with the entry into effect of Commission Regulation 1/2003 OJ 2003 L1/1.

82 See e.g. the handbook of the Financial Services Authority at www.fsa.gov.uk.

83 See McBarnet & Whelan (n 30).

84 See D McBarnet, "Whiter than white collar crime: tax, fraud insurance and the management of stigma" (1991) 42 *British Journal of Sociology* 323 for a discussion of the role of counsel's opinion in creating tax avoidance strategies.

85 See generally R McCormick, *Legal Risk* (n 5) 261-267.

86 Three instances of such uncertainty are discussed at F below.

in question even if not in the aggregate), then the pressure for uncertainty to be resolved is reduced and the possibility emerges of the law remaining uncertain without damaging consequences. There may also be a degree of path-dependence inherent in this process in the sense that a particular technique for resolving uncertainty may become so embedded in commercial practice that the underlying legal certainty which gave rise to it may be overlooked.

It is difficult to assess the extent to which this explanation is accurate. Part of the problem is the correct identification of when uncertainty exists in the first place. My approach, as indicated earlier, is to regard uncertainty as existing when, before any dispute or adjudication occurs, the parties to a transaction can be observed to have taken action that is indicative of the presence of legal uncertainty. On that view, legal uncertainty is not associated directly with reported cases in the courts. However, short of engaging in some empirical research, there is no set of recorded transactions against which my hypothesis can be tested other than reported cases. On that basis, I have selected three important instances of the persistence of long-term uncertainty in commercial law to illustrate my argument. The selection is of course biased towards my own interests and by the expectation that the cases will provide some backing for the hypothesis. However, these aspects of the cases are of much less significance than their two central characteristics, viz that they are examples of the persistence of long-term uncertainty in important commercial rules. So long as those two characteristics are present, the cases can be regarded as typical of the issue under examination. Of course, the extent to which any small set of cases can be regarded as supporting the hypothesis is open to question, but it is at least a start.

F. THREE EXAMPLES OF PERSISTENT UNCERTAINTY

The examples discussed below illustrate different aspects of uncertainty. The first is largely of historical interest while the other two are of more direct relevance to current law. The first and second relate to mandatory rules, in respect of which the effect of uncertainty is greatest because it is not possible to contract round such rules. The third refers to one of the major default rules in contract law. The main focus is twofold. The first is to identify why, in the light of the propositions made above, uncertainty persisted for so long in respect of basic rules of commercial law. None of the rules examined below can be regarded as trivial or as technicalities: they are fundamental within their field and therefore the presence of uncertainty should, in principle, carry significant implications. The second is to determine the effect of the response to uncertainty in each case and thereby to assess the extent to which a successful response is capable of perpetuating the existence of uncertainty.

(1) The nineteenth century law on corporate personality and limited liability

In *Salomon v Salomon*⁸⁷ uncertainty in relation to the requirements for a valid incorporation of a limited liability company facilitated the use of creative compliance to achieve an objective that was contrary to the spirit of the law.⁸⁸ The fundamental nature of that uncertainty is encapsulated by the opening words of the speech of Lord Halsbury LC:⁸⁹

My Lords, the important question in this case, I am not certain it is not the only question, is whether the respondent company was a company at all—whether in fact that artificial creation of the Legislature had been validly constituted in this instance.

In essence, uncertainty existed in respect of the requirements of the enabling rules in the Companies Act 1862 that would permit Salomon to form a company with a separate legal personality and to limit his liability as a shareholder in a manner that would not be available to a sole trader or a partnership. Section 6 of the Act provided that:

Any seven or more persons associated for any lawful purpose may, by subscribing their names to a memorandum of association, and otherwise complying with the requisitions of this Act in respect of registration, form an incorporated company with or without limited liability.

It was argued that the appellant, Aron Salomon, had not met these requirements because the seven members of the company that he had formed to take over the leather and shoe manufacturing business he had previously conducted as a sole trader consisted of himself, his wife, his daughter and four sons. It was claimed by the liquidator and Salomon's (unpaid) creditors that this resulted in the company being a mere "nominee and agent" of Salomon and that the company could not therefore have been properly incorporated in accordance with the Act.

Reversing the decision of the Court of Appeal,⁹⁰ the House of Lords decided that Salomon had satisfied the requirements of the Act. In so doing, the House of Lords recognised the possibility of the enabling rule of limited liability being made available to *de facto* one-person companies. This represented a significant development in the law of business organisations as, at this point in time, there was no recognition in the legal systems of the United Kingdom of a form of

87 [1897] AC 22.

88 The fact that the decision of the House of Lords upheld Salomon's conduct does not defeat the argument that *ex ante* he engaged in creative compliance. Had it been a case of straightforward compliance, it is unlikely that the case would ever have reached the House of Lords or gained its seminal status in company law textbooks.

89 [1897] AC 22 at 25.

90 [1895] 2 Ch 323.

business organisation that made limited liability available to a group of fewer than seven persons.

From the perspective of responses to uncertainty, *Salomon* had engaged in creative compliance in that there was formal compliance with the rule but evasion of the purpose of the minimum membership requirement, which was to delimit organisations that had the benefit of limited liability (larger businesses taking the corporate form) from those that did not (smaller businesses organised as partnerships and sole traders).⁹¹ *Salomon* legitimised the blurring of that distinction and contributed to its eventual abandonment. It is noteworthy that it was not until 35 years after the relevant Act took effect that this issue was tested and resolved.⁹² Why did it take so long for such a basic issue to be resolved?

One possible answer is that the minimum membership requirement was not significant for the structure of business organisations. That possibility is best viewed in terms of the implications of the requirement for the legal structure of a business. *Salomon* legitimised the concept of a limited liability company controlled by a single person. That issue had not been directly addressed by the Companies Act 1862 which, by focusing on membership rather than control, had left open the possibility of the law being tested so as to arrive at the conclusion that was reached in *Salomon*. By giving formal recognition to this type of company, the *Salomon* decision would seem to act as a catalyst for company formation, and while it is difficult to demonstrate causality in this respect, the sharply rising trend in company formation in the early years of the twentieth century lends credence to this view.⁹³ It seems likely therefore that the *Salomon* decision was significant for the structure of business organisations and that the presence of uncertainty had previously imposed some constraint on the choice of legal structure. The possibility of uncertainty persisting because of the insignificance of the issue can therefore be dismissed.

Another possible answer is that there had in fact been some testing of the rules relating to company formation prior to *Salomon*. For example, in *Re National*

91 The extent to which a minimum membership requirement served that purpose declined over time and had largely disappeared by 1907, when the Limited Partnership Act made limited liability available to partnerships. However, at the time of the *Salomon* decision in 1896, it remained the case that the only legal form in which limited liability was available was a company which met the minimum membership requirement of the Companies Acts.

92 The period is even longer if one regards the Companies Act 1862 as having continued the basic policy set by the Companies Act 1844 in distinguishing companies and partnerships by rules relating to the number of members.

93 See G Todd, "Some aspects of joint stock companies, 1844-1900" (1932) 4 *Economic History Review* 46; H A Shannon, "The limited companies of 1866-1883" (1933) 4 *Economic History Review* 290. Todd's figures show a rise in company registrations from 400 in 1864 to 1,250 in 1889, 3,200 in 1907 and 6,200 in 1914.

Debenture and Assets Corporation,⁹⁴ it was held that section 18 of the Companies Act 1862 did not prevent a court from investigating whether the required number of persons had signed the memorandum, and if they had not, from concluding that no valid company had been formed.⁹⁵ An earlier case had established the principles applicable for determining the extent to which fraud was present in an arrangement whereby overvalued assets were sold to a company in return for the issue of shares: it was held in *Re Ambrose Lake Tin and Copper Mining*⁹⁶ that, while such a scheme might have the effect of deceiving if an offer of shares were made to the public, it could not represent a fraud on a company which was fully aware of and consented to the terms.⁹⁷ While, however, those cases might be viewed as supporting elements of the decision in *Salomon*, they do not provide a reliable basis for predicting the case's outcome and cannot therefore be regarded as a convincing explanation for the persistence of uncertainty.⁹⁸

Finally, and most significantly, the formulation of the membership requirement in terms of numbers of persons rather than control left scope for disagreement over its essential purpose, and created scope for creative compliance. It is noteworthy that it was possible—in the sense of not being prohibited expressly by the Companies Act 1862 or the procedure for incorporation—for a company to be formed that was under the control of a single person.⁹⁹ Shannon, an economic historian, has summed up the position as being that “The one-man company was already (1875) in existence”.¹⁰⁰ It was not necessary to test the law to determine *ex ante* whether such a company could be formed because there was no express obstacle to formation that could be enforced by the Registrar of Companies when presented with an application for

94 [1891] 2 Ch 505.

95 Section 18 read: “A certificate of the incorporation of any company, given by the registrar, shall be conclusive evidence that all the requisitions of this Act in respect of registration have been complied with.”

96 (1880) LR 14 Ch D 390.

97 The central point, made by James LJ at 394, was that in a transaction whereby a (new) company issues the whole of its capital in return for a property, the value of the property is of limited relevance because the shares represent shares in the assets of the company (the property) and the only thing the company gives back to the vendors is (shares in) what they have given to the company. The position differs if the public are invited to buy shares as in this case the over-valuation may lead investors to believe that the company has assets and a business worth more than they really are.

98 It is relatively easy to make this point in hindsight but it must also have appeared so to Salomon's advisers for otherwise it seems unlikely that the case would have been taken to the House of Lords.

99 For general historical background on the Companies Acts and the requirements for incorporation, see P Ireland, “The rise of the limited liability company” (1984) 12 *International Journal of the Sociology of Law* 239.

100 Shannon (n 93) at 290 n 4. Shannon's conclusion was based on the observation of the number and nature of shareholdings and the names and addresses of members on the files at the Registrar of Companies.

registration. Rather, the onus lay on those, such as the liquidator and creditors in *Salomon v Salomon*, who asserted the negative hypothesis to make their case after a company had been formed.¹⁰¹ Moreover, there would be little incentive to challenge the legal structure so long as the company was a going concern which did not directly threaten the interests of creditors.¹⁰² Had it been otherwise – had, for example, the law been formulated so as to refer to control as well as number of members – it seems quite likely that the law would have been tested sooner. Shannon’s comment indicates that the one-person company was in demand twenty years before the decision in *Salomon*. Such demand would have created pressure either for eventual judicial acceptance of the one-person company or for legislative intervention.¹⁰³ In that sense, the sheer scale of creative compliance may be an important factor: if it indicates a solution that is superior through being more efficient, there may be little point in the courts resisting a development which would be reached in any event through the process of efficiency-driven evolution.¹⁰⁴

(2) The duty of disclosure in insurance

The presence of uncertainty in this instance is encapsulated by the opening words of Lord Mustill in *Pan Atlantic Insurance Co v Pine Top Insurance Co*:¹⁰⁵

My Lords, the two short questions upon which this appeal depends should, after more than 200 years of history, be capable of a ready answer. The controversy which they have aroused shows that they are not.

The two short questions referred to by Lord Mustill concerned the obligations that arise during the negotiations preceding the formation of a contract of insurance and were expressed in the following terms:¹⁰⁶

1. Where sections 18(2) and 20(2) of the [Marine Insurance] Act [1906] relate the test of materiality to a circumstance “which would influence the judgement of a prudent

101 As there were no express provisions in the Companies Act 1862 dealing with this issue, the argument would have to be based (as in *Salomon v Salomon*) on the supposed intentions of the Act.

102 The position changes following insolvency when limited liability poses a threat to the creditors.

103 While the scale of demand for the incorporation of one-person-controlled companies indicated the benefits to members, an overall view of efficiency would need to take into account the costs for creditors. The subsequent development of company law indicates acceptance that the former outweighs the latter, at least when protective measures for creditors are put in place.

104 See B.(5) above.

105 [1995] 1 AC 501 at 518. This case fits the proxy for uncertainty adopted by Kress (n 18) as, in respect of the “materiality” issue, it was a 3-2 majority decision in the House of Lords. The decision of the Court of Appeal ([1993] 1 Lloyd’s Rep 496) was unanimous although Nicholls VC expressed “unease at the outcome”. Leave to appeal was refused by the Court of Appeal but allowed by the House of Lords.

106 At 521 per Lord Mustill.

underwriter in fixing the premium, or determining whether he will take the risk”, must it be shown that full and accurate disclosure would have led the prudent underwriter to a different decision on accepting or rating the risk; or is a lesser standard of impact on the mind of the prudent underwriter sufficient; and, if so, what is that lesser standard? 2. Is the establishment of a material misrepresentation or non-disclosure sufficient to enable the underwriter to avoid the policy; or is it also necessary that the misrepresentation or non-disclosure has induced the making of the policy, either at all or on the terms on which it was made? If the latter, where lies the burden of proof?

The earlier decision of the Court of Appeal in *CTI v Oceanus*¹⁰⁷ had illustrated that the law relating to these issues was far from clear.¹⁰⁸ In that case, the court held that a fact was material for the purposes of sections 18(2) and 20(2) of the Marine Insurance Act 1906 if it was something that a prudent insurer would want to know when assessing a risk. There was no need to show that the relevant fact had a decisive influence on the decision either as to accepting the risk or setting the terms on which it was accepted. Non-disclosure or misrepresentation of a material fact, so defined, gave the insurer the right to avoid the policy. Whether correctly decided or not,¹⁰⁹ *CTI v Oceanus* was certainly controversial. It was referred to in *Pan Atlantic* in the Court of Appeal as “a remarkably unpopular decision not only in the legal profession but also in the insurance markets”.¹¹⁰ The main complaints made against it were that it set the test of materiality too low, with the result that the law was too harsh to the insured, and that allowing a defence of misrepresentation or non-disclosure to succeed even if the actual underwriter’s mind was unaffected was contrary to common sense and justice.¹¹¹

In *Pan Atlantic*, the House of Lords took a different approach. It followed the interpretation of material circumstances in *CTI v Oceanus* in so far as that interpretation rejected the need for a decisive influence on the mind of a prudent insurer. However, the House of Lords added a further requirement, derived from the general law of contract: before an underwriter could avoid a contract for non-disclosure of a material circumstance he had to show that he had been induced by the non-disclosure to enter into the policy on the relevant terms.¹¹²

107 *Container Transport International Inc v Oceanus Mutual Underwriting Association (Bermuda) Ltd (No 1)* [1984] 1 Lloyd’s Rep 476.

108 For the inherent difficulty of judging when the law is or is not clear, see [1995] 1 AC 501 at 514 per Lord Templeman.

109 It became clear later that it was, at least in part, wrongly decided, as it was reversed in part by the House of Lords decision in *Pan Atlantic* [1995] 1 AC 501.

110 [1995] 1 AC 501 at 528 per Lord Mustill.

111 The main criticisms of *CTI v Oceanus* were summarised by Lord Mustill in *Pan Atlantic* at 528.

112 Later cases have made clear that there may be instances in which the materiality of the undisclosed matters is so obvious that the court may presume that the underwriter was induced, but this is an evidential presumption and may be rebutted: see *St Paul Fire & Marine Ins Co v McConnell Dowell Constructors* [1995] 2 Lloyd’s Rep 116; *Gaelic Assignments Ltd v Sharp* 2001 SLT 914.

From the perspective of legal certainty, what is most interesting about the decisions in *Pan Atlantic* and *CTI v Oceanus* is that such controversy could exist in respect of an issue that had been litigated for 200 years and had been the subject of codification (in the Marine Insurance Act) almost 80 years earlier. Non-disclosure and misrepresentation are core issues in insurance law because they define the obligations of the parties during contractual negotiations and allow the insurer to be released from liability if those obligations are not fulfilled. Moreover, the international importance of the London-based insurance market would suggest that such basic issues should have been tested and resolved long before the penultimate decade of the twentieth century. How can such prolonged uncertainty be explained?

Several explanations drawn from the reasoning adopted by the House of Lords in *Pan Atlantic* can be offered. First, the codification of the common law relating to marine insurance in the Marine Insurance Act 1906 resulted in some uncertainty as to the relationship between the codified law and the pre-existing common law. The 1906 Act was not a complete codification, and section 91(2) provided that: “The rules of the common law including the law merchant, save in so far as they are inconsistent with the express provisions of this Act, shall continue to apply to marine insurance.” The Act therefore contained within it the possibility that issues relating to the common law would have to be re-visited at some point in the future. Both the materiality and the inducement issues that arose in *Pan Atlantic* fell into this category.

In respect of the materiality issue Lord Mustill noted that Lord Mansfield’s formulation of the duty of disclosure in *Carter v Boehm*,¹¹³ the first extended exposition of the doctrine, did not use the word “influence” or refer to the prudent underwriter.¹¹⁴ As regards the development of these aspects of the duty of disclosure, Lord Mustill went on to say that:¹¹⁵

As regards the test of influence on the mind I feel little doubt that the origins lay in the writers of the texts. These were a far more potent source of general principle than the scattered decisions of the courts.

Having surveyed the relevant texts,¹¹⁶ he concluded that:¹¹⁷

they furnish substantial support for the view that the duty of disclosure extended to all matters which would have been taken into account by the underwriter when assessing the risk (i.e. the “speculation”) which he was consenting to assume. That is in my opinion what the Act was intending to convey, and what it actually says.

113 (1766) 3 Burr 1905.

114 [1995] 1 AC 501 at 536.

115 At 537.

116 Note the emphasis placed on texts rather than cases.

117 At 538.

Codification also raised problems as regards the role of inducement in both non-disclosure and misrepresentation. Neither section 18(1) of the 1906 Act, which sets out the duty of disclosure, nor section 20(1), which provides for the consequences of a misrepresentation, refers expressly to a requirement that the particular insurer should have been induced into the contract by the non-disclosure or misrepresentation. The omission appears odd when it is borne in mind that, under the general law of contract, the right to rescind for misrepresentation requires the person to whom the misrepresentation is made to have relied on it.¹¹⁸ However, the saving provision in section 91(2) opened up the possibility that the Act should be interpreted as preserving the common law requirement for inducement, and that was how the House of Lords ultimately disposed of the matter.

Nevertheless, while the partial nature of the Act's codification, coupled with the presence of a saving provision, created an environment in which some legal uncertainty could survive, it does not answer the question why uncertainty persisted for such a long time. To that question, two possible answers can be given.

One is that insurers may have been reluctant, for reputational reasons, to exercise their legal rights so as to avoid liability other than in instances in which non-disclosure was blatant or amounted to fraud. Moreover, if, as was suggested by the expert evidence given in *CTI v Oceanus*,¹¹⁹ the general understanding in the market (among brokers as well as underwriters) was that an insurer could avoid liability for non-disclosure on the basis of information that was not decisive, there may have been an acknowledgment among insurers that the law was favourably inclined towards their interests and that its strict application might lead to calls for it to be changed. Such an explanation is of course speculative, but is not so far removed from the evolution of consumer insurance law – which saw insurers abandon their strict legal rights through self-regulation¹²⁰ – as to make it fanciful.

A second and perhaps more important explanation is the possibility that other mechanisms were devised to resolve the uncertainty. One mechanism that can be regarded as fulfilling this role is the warranty: the insured confirms a particular state of facts or undertakes that some particular things shall be done or not be

118 See G H Treitel, *The Law of Contract*, 12th edn, by E Peel (2007) 369 and, in respect of Scotland, W W McBryde, *The Law of Contract in Scotland*, 3rd edn (2007) 417.

119 [1984] 1 Lloyd's Rep 476 at 492-493 per Kerr LJ.

120 The (self-regulatory) Statement of General Insurance Practice issued by the Association of British Insurers has now been replaced by Financial Services Authority, *Insurance: Conduct of Business* ("ICOB") 7.3.6, which gives statutory effect to the content of that Statement. Despite this change, the point made about legal evolution through self-regulation remains valid.

done or that some condition shall be fulfilled.¹²¹ A warranty must be exactly complied with, whether it be material to the risk or not, and failure to do so discharges the insurer from liability as from the date of the breach.¹²² Warranties may be express or implied and may also be created by a “basis of contract” clause, which has the effect of transforming the contents of a proposal form into a warranty.¹²³ The significance of this transformation is that the information contained in the proposal form becomes subject to the law on warranties as well as the law relating to non-disclosure and misrepresentation. This avoids any dispute over issues such as materiality and inducement because the law on warranties requires strict compliance. The use of basis clauses and warranties can therefore be viewed as a contractual mechanism that resolves some of the uncertainty in the law relating to non-disclosure and misrepresentation. To the extent that they avoid rules which are uncertain, they can be regarded as having contributed to the persistence of that uncertainty.

It may be asked why a “basis of contract” clause approach did not operate to resolve the legal uncertainty that arose in either *CTI v Oceanus* or *Pan Atlantic*. In fact, the clause was not used in either case, perhaps because the large volume of information made available to the underwriters meant that the clause would have imposed an unfair burden on the insured and his broker. Both cases thus differ from simpler instances of commercial insurance proposals made on an insurer’s standard form, as exemplified by *Dawsons Ltd v Bonnin*.¹²⁴ On the assumption that *CTI v Oceanus* and *Pan Atlantic* are representative of practice in the London insurance market, the strength of the argument about the warranty as a mechanism for the resolution of legal uncertainty declines. Nonetheless, the use of “basis of contract” clauses remained significant in standardised commercial insurance and even in consumer insurance prior to self-regulatory control.¹²⁵ While it might be possible to argue that the purpose of such clauses was simply to expand the scope of the duty of disclosure, it can equally be argued that they served the purpose of resolving uncertainty in the law.

(3) Material breach in contract law

The Scots law on breach of contract provides an example of the survival of uncertainty as a result of the benefits associated with it. It also provides an

121 Marine Insurance Act 1906 s 33(1).

122 Marine Insurance Act 1906 s 33(3).

123 See *Dawsons Ltd v Bonnin* [1922] AC 413 for an example of a “basis of contract” clause. While such clauses are prohibited in consumer insurance by ICOB (n 120) 7.3.6, they can still be used in commercial insurance.

124 [1922] AC 413.

125 See n 120 above.

example of how certainty is linked to the formulation of the law in terms of principles and rules.

Only material breach justifies rescission.¹²⁶ Less serious breaches may give rise to a right in damages or a claim of retention or lien, but do not allow the innocent party to rescind the contract.¹²⁷ Moreover, an unjustified act of rescission may constitute a repudiation of the contract.¹²⁸ Following repudiation the innocent party has the option to rescind the contract. It follows that the definition of a material breach performs a crucial function in delimiting the respective rights and remedies of contracting parties.

The classic definition of material breach is that given by Lord President Dunedin in *Wade v Waldon*:¹²⁹

It is familiar law, and quite well settled by decision that in any contract which contains multifarious stipulations there are some which go so to the root of the contract that a breach of those stipulations entitles the party pleading the breach to declare that the contract is at an end. There are others which do not go to the root of the contract, but which are part of the contract, and which would give rise if broken to an action in damages.

Gloag interprets this passage to mean that the identification of material terms is “a pure question of construction whether, looking to the whole terms of the contract, one particular provision is or is not material”.¹³⁰ McBryde adopts a different view, proposing that whether or not a breach is material is primarily “a question of fact” and that “It is not to be decided solely by looking within the four corners of the contract”.¹³¹ More importantly from the perspective of legal certainty, neither regards the case law that this issue has generated as giving rise to clear guidance as to when a breach will or will not be considered material. MacQueen and Thomson indeed focus on the uncertainty faced by contracting parties in determining whether a breach is material, and on the contractual solutions that can be adopted in order to limit uncertainty.¹³²

Bearing in mind the test identified earlier as an indicator of uncertainty – the extent to which a legal response to uncertainty could be observed – several observations can be made about material breach. The first is that the courts have

126 See generally W M Gloag, *The Law of Contract*, 2nd edn (1929) 603; McBryde, *Contract* (n 118) 575.

127 This point is not entirely free from doubt: for a discussion of the background and support for the view put forward here, see McBryde, *Contract* (n 118) 562.

128 *Wade v Waldon* 1909 SC 571.

129 1909 SC 571 at 576.

130 Gloag, *Contract* (n 126) 603.

131 McBryde, *Contract* (n 118) 577. The author refers to the need to examine issues such as the extent of any defect or delay in performance.

132 H L MacQueen and J Thomson, *Contract Law in Scotland*, 2nd edn (2007) 209.

not attempted to limit the freedom of contracting parties to define what type of breach will justify rescission. In *Bell Bros (HP) v Aitken*¹³³ it was said that “parties can always agree in a contract to treat any breach as justifying rescission”.¹³⁴ The law on material breach thus represents a weak default rule which, at least in the commercial context, can be varied in whatever manner the parties agree. That aspect may also help to explain the relatively limited degree to which the law on material breach has given rise to litigation: if uncertainty in the default rule is routinely resolved by contractual provisions, it can be expected that litigation will thereby be avoided.

Secondly, the case law indicates that, while appropriate contractual provision may be sought as a means of limiting uncertainty, that objective is not always achieved. Several cases can be cited in which a contract term purporting to define material breach and control its consequences proved problematic. For example, in *Charisma Properties Ltd v Grayling (1944) Ltd*¹³⁵ a contract for the sale of two farms contained a clause that attempted to control the exercise by the seller of the remedy of rescission in the event of delay in payment of the purchase price.¹³⁶ While the construction that the seller had argued for was ultimately adopted by the court when the case was heard on appeal, it was noted by the court that it was only possible to reach such an outcome by regarding the addition of the word “prior” (before “written notice” in the contract term) as being mere surplusage. In *L Schuler AG v Wickman Machine Tool Sales Ltd*¹³⁷ the House of Lords held that reference in a contract governed by English law to a term as a “condition” did not in the particular circumstances give the other party the right to rescind the contract. The significance of the decision was that, as pointed out in the dissenting judgment of Lord Wilberforce, it had previously been assumed that the use of the description “condition” in respect of a contractual stipulation was in itself sufficient to indicate that the parties should have the right to terminate following breach. In *Schuler*, clause 11 of the contract gave the right to terminate following “material breach” but did not define what that meant, leaving the court to determine whether a condition could be construed as

133 1939 SC 577.

134 In the modern context, of course, that conclusion needs to take account of the controls on terms defining breach that are contained in the Unfair Contract Terms Act 1977 and the Unfair Terms in Consumer Contracts Regulations 1999, SI 1999/2083.

135 1996 SC 556.

136 The relevant term provided that “In the event that the said purchase price is not paid in full within 21 days of the date of entry, our clients shall be entitled to treat your clients as being in material breach of contract and to rescind the missives on giving prior written notice to that effect to your clients...”

137 [1974] AC 235.

providing a right to terminate independently of clause 11. Thus, even when the parties make express contractual provision to limit uncertainty, poor drafting may stand in the way of a successful outcome. *Schuler* also exposed a disagreement between the Court of Appeal and the House of Lords as to the admissibility of subsequent conduct in the construction of contracts, and so a further ground of uncertainty.¹³⁸ There are thus two levels of uncertainty: the first in the default rule of material breach and the second in the rules of construction which are applied when interpreting contractual provisions that modify the default rule.¹³⁹ While in any given contractual dispute there can only be one level of uncertainty (as the default rule is either applied or modified),¹⁴⁰ the problem for contracting parties arises *ex ante* in the initial decision whether to accept or modify the default rule on material breach.

It can be argued that uncertainty in the law on material breach has a positive aspect in that it directs the parties to determine what is material for their own purposes. Uncertainty is thus a catalyst for the creation of enhanced certainty for the contract. When the parties stipulate what they mean by material breach, legal risk and uncertainty is reduced. Of course, in order for that process to operate effectively, it has to be assumed that the parties or their advisers have a sufficient understanding of contract law. Whether that is always the case in the commercial world is open to debate, but there is good reason to believe that the standard of knowledge at which this process can operate is quite low.¹⁴¹ Because the focus of the law on material breach is on fundamental obligations, it corresponds with those matters that are likely to have been the subject of detailed negotiation and express agreement. One does not need to be a lawyer to understand the fundamentals of a commercial transaction: indeed, in many commercial contexts the process of contracting has evolved so that such terms are the subject of negotiation and the remainder are standard or “boiler plate” terms. Parties may regulate material breach without even analysing or discussing the matter in those terms: in other words, regulation of material breach arises as a by-product of the normal negotiating procedure.

138 Compare [1972] 2 All ER 1173, where the Court of Appeal was divided, with [1974] AC 235 at 252 per Lord Reid, 263 per Lord Simon of Glaisdale, where criticism was expressed of having regard to subsequent conduct.

139 See A Barron, “The rise of legal pragmatism in English contract law” (2006) 22 *Journal of Contract Law* 187.

140 The possibility of setting aside the default rule in its entirety is disregarded as there will always be a rule governing material breach in some form: either the default rule or a modified form of the default rule.

141 Although it might be argued, on the basis of the contract that was the subject of litigation in *Schuler*, that even large companies with the benefit of legal advice may fail to define what is meant by material breach.

Of course, if parties fail to make their own provision, matters will be governed by the default rule on material breach. As argued earlier,¹⁴² the main effect of uncertainty for default rules is that they cannot properly fulfil their function as a gap-filler and a mechanism for economising on transaction costs because uncertainty drives parties to contract round the rule. That observation prompts the question whether the default rule on material breach can be regarded as appropriate or efficient.

In the formulation of default rules the standard view is that they should be majoritarian: that is, they should take the form that the parties would have agreed had they considered the matter.¹⁴³ However, it is recognised that, in some circumstances, alternative formulations may be preferable. An example is the rule governing the award of damages for breach of contract established in *Hadley v Baxendale*,¹⁴⁴ which limits damages to those that are foreseeable or explicitly brought to the attention of the other side at the time of the making of the contract. This has been characterised as a “penalty” default rule because it forces a party which faces loss that is not foreseeable to reveal that information to the other side or face the prospect of non-recovery.¹⁴⁵ It is regarded as an efficient rule because it encourages parties facing unforeseeable losses to speak out, thereby allowing contract prices to adjust to the risks that arise from breach: in economic terms, it creates a “separating equilibrium” in which different types of contracting parties identify their special characteristics and are sorted through the competitive process in different groups at different prices.

The rule relating to material breach can be viewed as an efficient default rule, despite its tendency to encourage contracting out, because it promotes the achievement of legal (contractual) certainty, and the *ex ante* cost of achieving such certainty is likely to be less than the risks arising from *ex post* misjudgements due to uncertainty over the right to rescind or to claim damages. Its utility can be judged by reference to the other options that might be adopted, which are (a) a more precise default rule and (b) a mandatory rule. Although option (a) would have the effect of adjusting the scope of the default rule, there could be no guarantee that, in aggregate, its take-up would increase as a result of greater certainty. Moreover, greater certainty might result in some cases that could be considered to fall within the outer limits of the old (uncertain) rule no longer falling within the scope of the new (more certain) rule. Option (b) would be incompatible with the principle of freedom of contract because it would remove

142 See the discussion of rule types in B.(3) above.

143 See e.g. Ayres & Gertner (n 38).

144 (1854) 9 Ex 341. The same rule governs the award of damages in Scots contract law.

145 By Ayres & Gertner (n 38) at 95.

the freedom of the parties to stipulate what they consider to be their fundamental obligations.

A different way of expressing the same conclusion would be to say that the law on material breach is more appropriately expressed as a principle rather than in the form of detailed rules. It seems clear from the analysis above that more detailed rules would not produce greater legal certainty. This conclusion lends support to Braithwaite's argument¹⁴⁶ that a policy of favouring rules over principles in pursuit of legal certainty, as proposed by Raz,¹⁴⁷ may not always yield benefits.

G. CONCLUSIONS

This article asks why uncertainty persists over time in commercial law given the traditional emphasis on the need for legal certainty to facilitate business transactions. The main answer is the availability of solutions which remove or limit uncertainty in a manner that displaces the need for the law to evolve or change so as to remove the uncertainty. Those solutions are in the main contractual but there are also important techniques – primarily creative compliance and regulatory guidance – that do not involve contracting and therefore offer a mechanism for the resolution of uncertainty in relation to regulatory rules. The field of application of these solutions can be regarded as establishing a limit for legal uncertainty, which is the level of uncertainty that can be tolerated by commercial practice without the need for the law to change. That limit may be established at different levels in different areas of law and for different reasons, as illustrated by the three examples discussed in the previous section.

In the case of company formation, uncertainty survived because creative compliance facilitated an objective which was contrary to the spirit of the law. Commercial practice in the field of insurance was able to deal with a substantial degree of uncertainty in the law relating to disclosure partly because there was an appropriate contractual solution in the form of a warranty. By way of contrast, the very presence of uncertainty in the law on material breach of contract can be regarded as having contributed to its survival because it directed the contracting parties to a more certain (and efficient) solution. In the case of insurance, it was the strength of the contractual solution that led to the survival of uncertainty in the underlying rule, whereas in contract law the uncertainty in the underlying

¹⁴⁶ See B.(2) above.

¹⁴⁷ Raz (n 25).

rule was the *sine qua non* of a contractual solution that was more efficient even than a more certain underlying legal rule (defining material breach).

It follows that the tolerable limit for legal uncertainty is itself a function of the tools that the law provides to solve uncertainty. When it is possible for the full range of responses to uncertainty to be deployed, it is likely that commercial practice can tolerate greater legal uncertainty than when that is not the case. The conditions under which the full range of responses is available in any legal system can be said to be when freedom of contract is prioritised in contract law (thereby facilitating contractual solutions), and also when the contracting parties have access to informed legal advice (thereby facilitating both creative compliance and contractual solutions). The extent to which these two conditions are met will vary and so too will the costs and benefits of a response to uncertainty in any particular circumstances. Where, for example, the degree of uncertainty is low and the response is costly, it may be better to take the risk associated with uncertainty. On the other hand, when uncertainty may be present even to a small degree in many thousands or even millions of routine transactions, it may be better to avoid the risk.

The significance of legal uncertainty in commercial law can therefore be said to be quite low so long as an adequate response is available. This is not to say that legislators and courts should not strive for legal certainty, for plainly they should; but when they cannot achieve it, the result may not be as disastrous as is sometimes claimed.